

REVISED  
VERSION\*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>5</sup> : <b>A61K 38/00, C07K 1/00, 16/00, C12Q 1/68, C12P 19/34, C07H 21/02, 21/04</b>		A2	(11) International Publication Number: <b>WO 99/42118</b>
(21) International Application Number: PCT US99 03265		(43) International Publication Date: 26 August 1999 (26.08.99)	
(22) International Filing Date: 17 February 1999 (17.02.99)		(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).	
(30) Priority Data: 09/024,753 18 February 1998 (18.02.98) US 09/072,596 5 May 1998 (05.05.98) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicant: CORIXA CORPORATION (US US); 1-24 Colum- bia Street, Seattle, WA 98104 (US).		Published <i>With declaration under Article 17(2)(a)</i> <i>Without abstract, title not checked by the International Searching Authority</i>	
(72) Inventors: REED, Steven, G.; 2843 122nd Place N.E., Belle- vue, WA 98005 (US). SKEIKY, Yasar, A.; 8327 25th Street, Seattle, WA 98107 (US). DILLON, Davin, C., 21607 N.E. 24th Street, Redmond, WA 98053 (US). CAM- POS-NETO, Antonio; 9308 Midship Court N.E., Bain- bridge Island, WA 98110 (US). HOUGHTON, Ray- mond; 2636 242nd Place S.E., Bothell, WA 98021 (US). VEDVICK, Thomas, S.; 124 South 300th Place, Federal Way, WA 98003 (US). TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US). LODES, Michael, J.; 9223 36th Avenue S.W., Seattle, WA 98126 (US). HENDRICKSON, Ronald, C.; 4114 S.W. Charlestown Street, Seattle, WA 98116 (US).			

(54) Title: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT

AF	Algeria	ES	Spain	IS	Iceland	SI	Slovenia
AM	Armenia	ET	Ethiopia	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AX	Aland Islands	GA	Gabon	LV	Latvia	SZ	Switzerland
AZ	Azerbaijan	GB	Great Britain	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Moldova	TG	Togo
BB	Barbados	GU	Guam	MG	Madagascar	TH	Thailand
BE	Belgium	GN	Guinea	MK	Macedonia	TM	Turkmenistan
BF	Burkina Faso	GP	Guadeloupe	ML	Mali	TR	Turkey
BG	Bulgaria	HR	Croatia	MT	Malta	TT	Trinidad and Tobago
BI	Burundi	IE	Ireland	MX	Mexico	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BS	Bahamas	IS	Israel	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MY	Malaysia	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NP	Nepal	VN	Vietnam
CG	Congo	KR	South Korea	NI	Nicaragua	YE	Yemen
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Ivory Coast	KZ	Kazakhstan	NZ	New Zealand		
CM	Cameroun						

REVISED  
VERSION

PATENT COOPERATION TREATY

PCT

DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT  
(PCT Article 17(2)(a) and Rule 39)

Applicant's or agent's file reference 9532-023-228	IMPORTANT DECLARATION	Date of mailing (day/month/year) <b>22 JUNE 1999</b>
International application No. PCT/US99/03265	International filing date (day/month/year) 17 FEBRUARY 1999	(Earliest) Priority Date (day/month/year) 18 FEBRUARY 1998
International Patent Classification (IPC) or both national classification and IPC Please See Continuation Sheet.		
Applicant CORIXA CORPORATION		

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below.

1. ☐ The subject matter of the international application relates to:
- a. ☐ scientific theories.
  - b. ☐ mathematical theories.
  - c. ☐ plant varieties.
  - d. ☐ animal varieties.
  - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
  - f. ☐ schemes, rules or methods of doing business.
  - g. ☐ schemes, rules or methods of performing purely mental acts.
  - h. ☐ schemes, rules or methods of playing games.
  - i. ☐ methods for treatment of the human body by surgery or therapy.
  - j. ☐ methods for treatment of the animal body by surgery or therapy.
  - k. ☐ diagnostic methods practiced on the human or animal body.
  - l. ☐ mere presentations of information.
  - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
- ☐ the description
  - ☐ the claims
  - ☐ the drawings
3. ☒ The failure of the nucleotide and/or amino acid sequence listing to comply with the prescribed requirements prevents a meaningful search from being carried out:
- ☒ it does not comply with the prescribed standard
  - ☐ it is not in the prescribed machine readable form
4. Other reasons:

**DECLARATION OF NON-ESTABLISHMENT OF  
INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US99/03265

The International Patent Classification (IPC) or National Classification and IPC are as listed below:

IPC(5): A61K 38/00; C07K 1/00; C07K 16/00; C12Q 1/68; C12P 19/34; C07H 21/02, 21/04  
US Cl. 530/300, 350, 387.1; 435/6, 91.1, 91.2; 536/23.1, 24.3, 24.31, 24.32, 24.33

to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- $\gamma$  stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection see Chan and Kaufmann, in  
5 *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved diagnostic methods for detecting tuberculosis. The present invention fulfills this need and further provides other related advantages.

#### SUMMARY OF THE INVENTION

Briefly stated, the present invention provides compositions and methods for diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of such an antigen  
15 that differs only in conservative substitutions and/or modifications. In one embodiment of this aspect, the soluble antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);
- (d) Thr-Lys-Ile-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Ile-Gly-Pro (SEQ ID NO: 118);
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID

- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123);
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131)

wherein Xaa may be any amino acid.

In a related aspect, polypeptides are provided comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, the antigen having one of the following N-terminal sequences:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 122) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

substitutions and/or modifications, wherein the antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178, 184-188, 194-196, 198, 210-220, 232, 234, 235, 237-242, 248-251, 256-271, 287, 288, 290-293 and 298-337, ,  
5 the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178, 184-188, 194-196, 198, 210-220, 232, 234, 235, 237-242, 248-251, 256-271, 287, 288, 290-293 and 298-337, or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides,  
10 recombinant expression vectors comprising these DNA sequences and host cells transformed or transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

15 In further aspects of the subject invention, methods and diagnostic kits are provided for detecting tuberculosis in a patient. The methods comprise:  
(a) contacting a biological sample with at least one of the above polypeptides; and  
(b) detecting in the sample the presence of antibodies that bind to the polypeptide or polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.  
20 Suitable biological samples include whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine. The diagnostic kits comprise one or more of the above polypeptides in combination with a detection reagent.

The present invention also provides methods for detecting *M. tuberculosis* infection comprising:  
(a) obtaining a biological sample from a patient;  
25 (b) contacting the sample with at least one oligonucleotide primer in a polymerase chain reaction, the oligonucleotide primer being specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers. In one embodiment, the

In a further aspect, the present invention provides a method for detecting *M. tuberculosis* infection in a patient comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with an oligonucleotide probe specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe. In one embodiment, the oligonucleotide probe comprises at least about 15 contiguous nucleotides of such a DNA sequence.

In yet another aspect, the present invention provides antibodies, both polyclonal and monoclonal, that bind to the polypeptides described above, as well as methods for their use in the detection of *M. tuberculosis* infection.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

#### BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

Figures 2A-D illustrate the reactivity of antisera raised against secretory *M. tuberculosis* proteins, the known *M. tuberculosis* antigen 85b and the inventive antigens Tb38-1 and TbH-9, respectively, with *M. tuberculosis* lysate (lane 2), *M. tuberculosis* secretory proteins (lane 3), recombinant Tb38-1 (lane 4), recombinant TbH-9 (lane 5) and recombinant 85b (lane 6).

Figure 3A illustrates the stimulation of proliferation in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, recombinant TbH-9 and a control antigen, TbRa11.

Figure 3B illustrates the stimulation of interferon- $\gamma$  production in a TbH-

Figure 4 illustrates the reactivity of two representative polypeptides with sera from *M. tuberculosis*-infected and uninfected individuals, as compared to the reactivity of bacterial lysate.

Figure 5 shows the reactivity of four representative polypeptides with sera from *M. tuberculosis*-infected and uninfected individuals, as compared to the reactivity of the 38 kD antigen.

Figure 6 shows the reactivity of recombinant 38 kD and TbRa11 antigens with sera from *M. tuberculosis* patients, PPD positive donors and normal donors.

Figure 7 shows the reactivity of the antigen TbRa2A with 38 kD negative sera.

Figure 8 shows the reactivity of the antigen of SEQ ID NO: 60 with sera from *M. tuberculosis* patients and normal donors.

Figure 9 illustrates the reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with sera from *M. tuberculosis* patients, PPD positive donors and normal donors as determined by indirect ELISA.

Figure 10 illustrates the reactivity of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from *M. tuberculosis* patients and from normal donors, and with a pool of sera from *M. tuberculosis* patients, as determined both by direct and indirect ELISA.

Figure 11 illustrates the reactivity of increasing concentrations of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from *M. tuberculosis* patients and from normal donors as determined by ELISA.

Figures 12A-E illustrate the reactivity of the recombinant antigens MO-1, MO-2, MO-4, MO-28 and MO-29, respectively, with sera from *M. tuberculosis* patients and from normal donors as determined by ELISA.

SEQ ID NO: 1 is the DNA sequence of TbRa1

- SEQ. ID NO. 4 is the DNA sequence of TbRa12.  
SEQ. ID NO. 5 is the DNA sequence of TbRa13.  
SEQ. ID NO. 6 is the DNA sequence of TbRa16.  
SEQ. ID NO. 7 is the DNA sequence of TbRa17.  
5 SEQ. ID NO. 8 is the DNA sequence of TbRa18.  
SEQ. ID NO. 9 is the DNA sequence of TbRa19.  
SEQ. ID NO. 10 is the DNA sequence of TbRa24.  
SEQ. ID NO. 11 is the DNA sequence of TbRa26.  
SEQ. ID NO. 12 is the DNA sequence of TbRa28.  
10 SEQ. ID NO. 13 is the DNA sequence of TbRa29.  
SEQ. ID NO. 14 is the DNA sequence of TbRa2A.  
SEQ. ID NO. 15 is the DNA sequence of TbRa3.  
SEQ. ID NO. 16 is the DNA sequence of TbRa32.  
SEQ. ID NO. 17 is the DNA sequence of TbRa35.  
15 SEQ. ID NO. 18 is the DNA sequence of TbRa36.  
SEQ. ID NO. 19 is the DNA sequence of TbRa4.  
SEQ. ID NO. 20 is the DNA sequence of TbRa9.  
SEQ. ID NO. 21 is the DNA sequence of TbRaB.  
SEQ. ID NO. 22 is the DNA sequence of TbRaC.  
20 SEQ. ID NO. 23 is the DNA sequence of TbRaD.  
SEQ. ID NO. 24 is the DNA sequence of YYWCPCG.  
SEQ. ID NO. 25 is the DNA sequence of AAMK.  
SEQ. ID NO. 26 is the DNA sequence of Tbl-23.  
SEQ. ID NO. 27 is the DNA sequence of Tbl-24.  
25 SEQ. ID NO. 28 is the DNA sequence of Tbl-25.  
SEQ. ID NO. 29 is the DNA sequence of Tbl-28.  
SEQ. ID NO. 30 is the DNA sequence of Tbl-29.  
SEQ. ID NO. 31 is the DNA sequence of Tbl-3.

SEQ. ID NO. 34 is the DNA sequence of TbM-1.

SEQ. ID NO. 35 is the DNA sequence of TbM-3.

SEQ. ID NO. 36 is the DNA sequence of TbM-6.

SEQ. ID NO. 37 is the DNA sequence of TbM-7.

5 SEQ. ID NO. 38 is the DNA sequence of TbM-9.

SEQ. ID NO. 39 is the DNA sequence of TbM-12.

SEQ. ID NO. 40 is the DNA sequence of TbM-13.

SEQ. ID NO. 41 is the DNA sequence of TbM-14.

SEQ. ID NO. 42 is the DNA sequence of TbM-15.

10 SEQ. ID NO. 43 is the DNA sequence of TbH-4.

SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.

SEQ. ID NO. 45 is the DNA sequence of TbH-12.

SEQ. ID NO. 46 is the DNA sequence of Tb38-1.

SEQ. ID NO. 47 is the DNA sequence of Tb38-4.

15 SEQ. ID NO. 48 is the DNA sequence of TbL-17.

SEQ. ID NO. 49 is the DNA sequence of TbL-20.

SEQ. ID NO. 50 is the DNA sequence of TbL-21.

SEQ. ID NO. 51 is the DNA sequence of TbH-16.

SEQ. ID NO. 52 is the DNA sequence of DPEP.

20 SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.

SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.

SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.

SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.

SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.

25 SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.

SEQ. ID NO. 59 is the protein sequence of AEES N-terminal Antigen.

SEQ. ID NO. 60 is the protein sequence of DPEP N-terminal Antigen.

SEQ. ID NO. 61 is the protein sequence of APKT N-terminal Antigen.

SEQ. ID NO. 64 is the deduced amino acid sequence of TbRa1.

SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa10.

SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa11.

SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa12.

5 SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa13.

SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa16.

SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa17.

SEQ. ID NO. 71 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 72 is the deduced amino acid sequence of TbRa19.

10 SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa24.

SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa26.

SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa28.

SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa29.

SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa2A.

15 SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa3.

SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa32.

SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa35.

SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa36.

SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa4.

20 SEQ. ID NO. 83 is the deduced amino acid sequence of TbRa9.

SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaB.

SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaC.

SEQ. ID NO. 86 is the deduced amino acid sequence of TbRaD.

SEQ. ID NO. 87 is the deduced amino acid sequence of YYWCPG.

25 SEQ. ID NO. 88 is the deduced amino acid sequence of TbAAMK.

SEQ. ID NO. 89 is the deduced amino acid sequence of Tb38-1.

SEQ. ID NO. 90 is the deduced amino acid sequence of TbH-4.

SEQ. ID NO. 91 is the deduced amino acid sequence of TbH-5.

SEQ. ID NO. 94 is the DNA sequence of DPAS.

SEQ. ID NO. 95 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 96 is the DNA sequence of DPV.

SEQ. ID NO. 97 is the deduced amino acid sequence of DPV.

5 SEQ. ID NO. 98 is the DNA sequence of ESAT-6.

SEQ. ID NO. 99 is the deduced amino acid sequence of ESAT-6.

SEQ. ID NO. 100 is the DNA sequence of TbH-8-2.

SEQ. ID NO. 101 is the DNA sequence of TbH-9FL.

SEQ. ID NO. 102 is the deduced amino acid sequence of TbH-9FL

10 SEQ. ID NO. 103 is the DNA sequence of TbH-9-1.

SEQ. ID NO. 104 is the deduced amino acid sequence of TbH-9-1.

SEQ. ID NO. 105 is the DNA sequence of TbH-9-4.

SEQ. ID NO. 106 is the deduced amino acid sequence of TbH-9-4.

SEQ. ID NO. 107 is the DNA sequence of Tb38-1F2 IN.

15 SEQ. ID NO. 108 is the DNA sequence of Tb38-1F2 RP.

SEQ. ID NO. 109 is the deduced amino acid sequence of Tb37-FL.

SEQ. ID NO. 110 is the deduced amino acid sequence of Tb38-IN.

SEQ. ID NO. 111 is the DNA sequence of Tb38-1F3.

SEQ. ID NO. 112 is the deduced amino acid sequence of Tb38-1F3.

20 SEQ. ID NO. 113 is the DNA sequence of Tb38-1F5.

SEQ. ID NO. 114 is the DNA sequence of Tb38-1F6.

SEQ. ID NO. 115 is the deduced N-terminal amino acid sequence of DPV.

SEQ. ID NO. 116 is the deduced N-terminal amino acid sequence of AVGS.

SEQ. ID NO. 117 is the deduced N-terminal amino acid sequence of AAMK.

25 SEQ. ID NO. 118 is the deduced N-terminal amino acid sequence of YYWC.

SEQ. ID NO. 119 is the deduced N-terminal amino acid sequence of DIGS.

SEQ. ID NO. 120 is the deduced N-terminal amino acid sequence of AAES.

SEQ. ID NO. 121 is the deduced N-terminal amino acid sequence of DPVP.

SEQ. ID NO. 124 is the protein sequence of DPPD N-terminal Antigen.

SEQ ID NO. 125-128 are the protein sequences of four DPPD cyanogen bromide fragments.

SEQ ID NO. 129 is the N-terminal protein sequence of XDS antigen.

5 SEQ ID NO. 130 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 131 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 132 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 133 is the DNA sequence of TbH-29.

SEQ ID NO. 134 is the DNA sequence of TbH-30.

10 SEQ ID NO. 135 is the DNA sequence of TbH-32.

SEQ ID NO. 136 is the DNA sequence of TbH-33.

SEQ ID NO. 137 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 138 is the predicted amino acid sequence of TbH-30.

SEQ ID NO. 139 is the predicted amino acid sequence of TbH-32.

15 SEQ ID NO. 140 is the predicted amino acid sequence of TbH-33.

SEQ ID NO. 141-146 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO. 147 is the DNA sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

20 SEQ ID NO. 148 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO. 149 is the DNA sequence of the M. tuberculosis antigen 38 kD.

SEQ ID NO. 150 is the amino acid sequence of the M. tuberculosis antigen 38 kD.

25 SEQ ID NO. 151 is the DNA sequence of XP14.

SEQ ID NO. 152 is the DNA sequence of XP24.

SEQ ID NO. 153 is the DNA sequence of XP31.

SEQ ID NO. 154 is the 5' DNA sequence of XP33.

SEQ ID NO: 157 is the predicted amino acid sequence encoded by the reverse complement of XP14.

SEQ ID NO: 158 is the DNA sequence of XP27.

SEQ ID NO: 159 is the DNA sequence of XP36.

5 SEQ ID NO: 160 is the 5' DNA sequence of XP4.

SEQ ID NO: 161 is the 5' DNA sequence of XP5.

SEQ ID NO: 162 is the 5' DNA sequence of XP17.

SEQ ID NO: 163 is the 5' DNA sequence of XP30.

SEQ ID NO: 164 is the 5' DNA sequence of XP2.

10 SEQ ID NO: 165 is the 3' DNA sequence of XP2.

SEQ ID NO: 166 is the 5' DNA sequence of XP3.

SEQ ID NO: 167 is the 3' DNA sequence of XP3.

SEQ ID NO: 168 is the 5' DNA sequence of XP6.

SEQ ID NO: 169 is the 3' DNA sequence of XP6.

15 SEQ ID NO: 170 is the 5' DNA sequence of XP18.

SEQ ID NO: 171 is the 3' DNA sequence of XP18.

SEQ ID NO: 172 is the 5' DNA sequence of XP19.

SEQ ID NO: 173 is the 3' DNA sequence of XP19.

SEQ ID NO: 174 is the 5' DNA sequence of XP22.

20 SEQ ID NO: 175 is the 3' DNA sequence of XP22.

SEQ ID NO: 176 is the 5' DNA sequence of XP25.

SEQ ID NO: 177 is the 3' DNA sequence of XP25.

SEQ ID NO: 178 is the full-length DNA sequence of TbH4-XP1.

SEQ ID NO: 179 is the predicted amino acid sequence of TbH4-XP1.

25 SEQ ID NO: 180 is the predicted amino acid sequence encoded by the reverse complement of TbH4-XP1.

SEQ ID NO: 181 is a first predicted amino acid sequence encoded by XP36.

SEQ ID NO: 182 is a second predicted amino acid sequence encoded by XP36.

SEQ ID NO: 184 is the DNA sequence of RDIF2.

SEQ ID NO: 185 is the DNA sequence of RDIF5.

SEQ ID NO: 186 is the DNA sequence of RDIF8.

SEQ ID NO: 187 is the DNA sequence of RDIF10.

5 SEQ ID NO: 188 is the DNA sequence of RDIF11.

SEQ ID NO: 189 is the predicted amino acid sequence of RDIF2.

SEQ ID NO: 190 is the predicted amino acid sequence of RDIF5.

SEQ ID NO: 191 is the predicted amino acid sequence of RDIF8.

SEQ ID NO: 192 is the predicted amino acid sequence of RDIF10.

10 SEQ ID NO: 193 is the predicted amino acid sequence of RDIF11.

SEQ ID NO: 194 is the 5' DNA sequence of RDIF12.

SEQ ID NO: 195 is the 3' DNA sequence of RDIF12.

SEQ ID NO: 196 is the DNA sequence of RDIF7.

SEQ ID NO: 197 is the predicted amino acid sequence of RDIF7.

15 SEQ ID NO: 198 is the DNA sequence of DIF2-1.

SEQ ID NO: 199 is the predicted amino acid sequence of DIF2-1.

SEQ ID NO: 200-207 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as TbF-2).

20 SEQ ID NO: 208 is the DNA sequence of the fusion protein TbF-2.

SEQ ID NO: 209 is the amino acid sequence of the fusion protein TbF-2.

SEQ ID NO: 210 is the 5' DNA sequence of MO-1.

SEQ ID NO: 211 is the 5' DNA sequence for MO-2.

SEQ ID NO: 212 is the 5' DNA sequence for MO-4.

25 SEQ ID NO: 213 is the 5' DNA sequence for MO-8.

SEQ ID NO: 214 is the 5' DNA sequence for MO-9.

SEQ ID NO: 215 is the 5' DNA sequence for MO-26.

SEQ ID NO: 216 is the 5' DNA sequence for MO-28.

SEQ ID NO: 219 is the 5' DNA sequence for MO-34.

SEQ ID NO: 220 is the 5' DNA sequence for MO-35.

SEQ ID NO: 221 is the predicted amino acid sequence for MO-1.

SEQ ID NO: 222 is the predicted amino acid sequence for MO-2.

5 SEQ ID NO: 223 is the predicted amino acid sequence for MO-4.

SEQ ID NO: 224 is the predicted amino acid sequence for MO-8.

SEQ ID NO: 225 is the predicted amino acid sequence for MO-9.

SEQ ID NO: 226 is the predicted amino acid sequence for MO-26.

SEQ ID NO: 227 is the predicted amino acid sequence for MO-28.

10 SEQ ID NO: 228 is the predicted amino acid sequence for MO-29.

SEQ ID NO: 229 is the predicted amino acid sequence for MO-30.

SEQ ID NO: 230 is the predicted amino acid sequence for MO-34.

SEQ ID NO: 231 is the predicted amino acid sequence for MO-35.

SEQ ID NO: 232 is the determined DNA sequence for MO-10.

15 SEQ ID NO: 233 is the predicted amino acid sequence for MO-10.

SEQ ID NO: 234 is the 3' DNA sequence for MO-27.

SEQ ID NO: 235 is the full-length DNA sequence for DPPD.

SEQ ID NO: 236 is the predicted full-length amino acid sequence for DPPD.

SEQ ID NO: 237 is the determined 5' cDNA sequence for LSER-10.

20 SEQ ID NO: 238 is the determined 5' cDNA sequence for LSER-11.

SEQ ID NO: 239 is the determined 5' cDNA sequence for LSER-12.

SEQ ID NO: 240 is the determined 5' cDNA sequence for LSER-13.

SEQ ID NO: 241 is the determined 5' cDNA sequence for LSER-16.

SEQ ID NO: 242 is the determined 5' cDNA sequence for LSER-25.

25 SEQ ID NO: 243 is the predicted amino acid sequence for LSER-10.

SEQ ID NO: 244 is the predicted amino acid sequence for LSER-12.

SEQ ID NO: 245 is the predicted amino acid sequence for LSER-13.

SEQ ID NO: 246 is the predicted amino acid sequence for LSER-16.

SEQ ID NO: 249 is the determined cDNA sequence for LSER-23  
SEQ ID NO: 250 is the determined cDNA sequence for LSER-24  
SEQ ID NO: 251 is the determined cDNA sequence for LSER-27  
SEQ ID NO: 252 is the predicted amino acid sequence for LSER-18  
5 SEQ ID NO: 253 is the predicted amino acid sequence for LSER-23  
SEQ ID NO: 254 is the predicted amino acid sequence for LSER-24  
SEQ ID NO: 255 is the predicted amino acid sequence for LSER-27  
SEQ ID NO: 256 is the determined 5' cDNA sequence for LSER-1  
SEQ ID NO: 257 is the determined 5' cDNA sequence for LSER-3  
10 SEQ ID NO: 258 is the determined 5' cDNA sequence for LSER-4  
SEQ ID NO: 259 is the determined 5' cDNA sequence for LSER-5  
SEQ ID NO: 260 is the determined 5' cDNA sequence for LSER-6  
SEQ ID NO: 261 is the determined 5' cDNA sequence for LSER-8  
SEQ ID NO: 262 is the determined 5' cDNA sequence for LSER-14  
15 SEQ ID NO: 263 is the determined 5' cDNA sequence for LSER-15  
SEQ ID NO: 264 is the determined 5' cDNA sequence for LSER-17  
SEQ ID NO: 265 is the determined 5' cDNA sequence for LSER-19  
SEQ ID NO: 266 is the determined 5' cDNA sequence for LSER-20  
SEQ ID NO: 267 is the determined 5' cDNA sequence for LSER-22  
20 SEQ ID NO: 268 is the determined 5' cDNA sequence for LSER-26  
SEQ ID NO: 269 is the determined 5' cDNA sequence for LSER-28  
SEQ ID NO: 270 is the determined 5' cDNA sequence for LSER-29  
SEQ ID NO: 271 is the determined 5' cDNA sequence for LSER-30  
SEQ ID NO: 272 is the predicted amino acid sequence for LSER-1  
25 SEQ ID NO: 273 is the predicted amino acid sequence for LSER-3  
SEQ ID NO: 274 is the predicted amino acid sequence for LSER-5  
SEQ ID NO: 275 is the predicted amino acid sequence for LSER-6  
SEQ ID NO: 276 is the predicted amino acid sequence for LSER-8

SEQ ID NO: 279 is the predicted amino acid sequence for LSER-17

SEQ ID NO: 280 is the predicted amino acid sequence for LSER-19

SEQ ID NO: 281 is the predicted amino acid sequence for LSER-20

SEQ ID NO: 282 is the predicted amino acid sequence for LSER-22

5 SEQ ID NO: 283 is the predicted amino acid sequence for LSER-26

SEQ ID NO: 284 is the predicted amino acid sequence for LSER-28

SEQ ID NO: 285 is the predicted amino acid sequence for LSER-29

SEQ ID NO: 286 is the predicted amino acid sequence for LSER-30

SEQ ID NO: 287 is the determined cDNA sequence for LSER-9

10 SEQ ID NO: 288 is the determined cDNA sequence for the reverse complement of LSER-6

SEQ ID NO: 289 is the predicted amino acid sequence for the reverse complement of LSER-6

SEQ ID NO: 290 is the determined 5' cDNA sequence for MO-12

15 SEQ ID NO: 291 is the determined 5' cDNA sequence for MO-13

SEQ ID NO: 292 is the determined 5' cDNA sequence for MO-19

SEQ ID NO: 293 is the determined 5' cDNA sequence for MO-39

SEQ ID NO: 294 is the predicted amino acid sequence for MO-12

SEQ ID NO: 295 is the predicted amino acid sequence for MO-13

20 SEQ ID NO: 296 is the predicted amino acid sequence for MO-19

SEQ ID NO: 297 is the predicted amino acid sequence for MO-39

SEQ ID NO: 298 is the determined 5' cDNA sequence for Erdsn-1

SEQ ID NO: 299 is the determined 5' cDNA sequence for Erdsn-2

SEQ ID NO: 300 is the determined 5' cDNA sequence for Erdsn-4

25 SEQ ID NO: 301 is the determined 5' cDNA sequence for Erdsn-5

SEQ ID NO: 302 is the determined 5' cDNA sequence for Erdsn-6

SEQ ID NO: 303 is the determined 5' cDNA sequence for Erdsn-7

SEQ ID NO: 304 is the determined 5' cDNA sequence for Erdsn-8

SEQ ID NO: 307 is the determined 5' cDNA sequence for Erdsn-12  
SEQ ID NO: 308 is the determined 5' cDNA sequence for Erdsn-13  
SEQ ID NO: 309 is the determined 5' cDNA sequence for Erdsn-14  
SEQ ID NO: 310 is the determined 5' cDNA sequence for Erdsn-15  
5 SEQ ID NO: 311 is the determined 5' cDNA sequence for Erdsn-16  
SEQ ID NO: 312 is the determined 5' cDNA sequence for Erdsn-17  
SEQ ID NO: 313 is the determined 5' cDNA sequence for Erdsn-18  
SEQ ID NO: 314 is the determined 5' cDNA sequence for Erdsn-21  
SEQ ID NO: 315 is the determined 5' cDNA sequence for Erdsn-22  
10 SEQ ID NO: 316 is the determined 5' cDNA sequence for Erdsn-23  
SEQ ID NO: 317 is the determined 5' cDNA sequence for Erdsn-25  
SEQ ID NO: 318 is the determined 3' cDNA sequence for Erdsn-1  
SEQ ID NO: 319 is the determined 3' cDNA sequence for Erdsn-2  
SEQ ID NO: 320 is the determined 3' cDNA sequence for Erdsn-4  
15 SEQ ID NO: 321 is the determined 3' cDNA sequence for Erdsn-5  
SEQ ID NO: 322 is the determined 3' cDNA sequence for Erdsn-7  
SEQ ID NO: 323 is the determined 3' cDNA sequence for Erdsn-8  
SEQ ID NO: 324 is the determined 3' cDNA sequence for Erdsn-9  
SEQ ID NO: 325 is the determined 3' cDNA sequence for Erdsn-10  
20 SEQ ID NO: 326 is the determined 3' cDNA sequence for Erdsn-12  
SEQ ID NO: 327 is the determined 3' cDNA sequence for Erdsn-13  
SEQ ID NO: 328 is the determined 3' cDNA sequence for Erdsn-14  
SEQ ID NO: 329 is the determined 3' cDNA sequence for Erdsn-15  
SEQ ID NO: 330 is the determined 3' cDNA sequence for Erdsn-16  
25 SEQ ID NO: 331 is the determined 3' cDNA sequence for Erdsn-17  
SEQ ID NO: 332 is the determined 3' cDNA sequence for Erdsn-18  
SEQ ID NO: 333 is the determined 3' cDNA sequence for Erdsn-21  
SEQ ID NO: 334 is the determined 3' cDNA sequence for Erdsn-22

SEQ ID NO: 337 is the determined cDNA sequence for Erdsn-24

SEQ ID NO: 338 is the determined amino acid sequence for a *M. tuberculosis*  
85b precursor homolog

SEQ ID NO: 339 is the determined amino acid sequence for spot 1

5 SEQ ID NO: 340 is a determined amino acid sequence for spot 2

SEQ ID NO: 341 is a determined amino acid sequence for spot 2

SEQ ID NO: 342 is the determined amino acid seq for spot 4

SEQ ID NO: 343 is the sequence of primer PDM-157

SEQ ID NO: 344 is the sequence of primer PDM-160

10 SEQ ID NO: 345 is the DNA sequence of the fusion protein TbF-6

SEQ ID NO: 346 is the amino acid sequence of fusion protein TbF-6

SEQ ID NO: 347 is the sequence of primer PDM-176

SEQ ID NO: 348 is the sequence of primer PDM-175

SEQ ID NO: 349 is the DNA sequence of the fusion protein TbF-8

15 SEQ ID NO: 350 is the amino acid sequence of the fusion protein TbF-8

#### DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to  
20 compositions and methods for diagnosing tuberculosis. The compositions of the subject  
invention include polypeptides that comprise at least one antigenic portion of a  
*M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative  
substitutions and/or modifications. Polypeptides within the scope of the present  
invention include, but are not limited to, soluble *M. tuberculosis* antigens. A "soluble  
25 *M. tuberculosis* antigen" is a protein of *M. tuberculosis* origin that is present in  
*M. tuberculosis* culture filtrate. As used herein, the term "polypeptide" encompasses  
amino acid chains of any length, including full length proteins (i.e., antigens), wherein  
the amino acid residues are linked by covalent bonds (e.g., peptide bonds).

be derived from the native *M. tuberculosis* antigen or may be heterologous, and such sequences may (but need not) be antigenic.

An "antigenic portion" of an antigen (which may or may not be soluble) is a portion that is capable of reacting with sera obtained from an *M. tuberculosis*-infected individual (i.e., generates an absorbance reading with sera from infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals, in a representative ELISA assay described herein). An "*M. tuberculosis*-infected individual" is a human who has been infected with *M. tuberculosis* (e.g., has an intradermal skin test response to PPD that is at least 0.5 cm in diameter). Infected individuals may display symptoms of tuberculosis or may be free of disease symptoms. Polypeptides comprising at least an antigenic portion of one or more *M. tuberculosis* antigens as described herein may generally be used, alone or in combination, to detect tuberculosis in a patient.

The compositions and methods of the present invention also encompass variants of the above polypeptides and DNA molecules. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic, antigenic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of tuberculosis. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

As used herein, a "conservative substitution" is one in which an amino

hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

5 Variants may also, or alternatively, contain other modifications, including the deletion or addition of amino acids that have minimal influence on the antigenic properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

A nucleotide "variant" is a sequence that differs from the recited nucleotide sequence in having one or more nucleotide deletions, substitutions or additions. Such modifications may be readily introduced using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis as taught, for example, by Adelman et al. (*DNA*, 2:183, 1983). Nucleotide variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variant nucleotide sequences preferably exhibit at least about 70%, more preferably at least about 80% and most preferably at least about 90% identity to the recited sequence. Such variant nucleotide sequences will generally hybridize to the recite nucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 0.5X SSC, 0.2% SDS, hybridizing at 65 °C, 0.5X SSC, 0.2% SDS overnight, followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65 °C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65 °C.

In a related aspect, combination, or fusion, polypeptides are disclosed. A fusion polypeptide is a polypeptide in which two or more polypeptides are joined together.

joined directly (i.e., with no intervening amino acids) or may be joined by way of a linker sequence (e.g., Gly-Cys-Gly) that does not significantly diminish the antigenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such  
5 antigens, may be prepared using any of a variety of procedures. For example, soluble antigens may be isolated from *M. tuberculosis* culture filtrate by procedures known to those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens may then be evaluated for a desired property, such as the ability to react with sera obtained from an *M. tuberculosis*-infected individual.  
10 Such screens may be performed using the representative methods described herein. Antigens may then be partially sequenced using, for example, traditional Edman chemistry. See Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and  
15 expressed in an appropriate host. DNA molecules encoding soluble antigens may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (e.g., rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may not be soluble may be identified by screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera  
20 obtained from patients infected with *M. tuberculosis*. Such screens may generally be performed using techniques well known in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989.

DNA sequences encoding soluble antigens may also be obtained by  
25 screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as

therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

5           Regardless of the method of preparation, the antigens described herein are "antigenic." More specifically, the antigens have the ability to react with sera obtained from an *M. tuberculosis*-infected individual. Reactivity may be evaluated using, for example, the representative ELISA assays described herein, where an absorbance reading with sera from infected individuals that is at least three standard  
10       deviations above the absorbance obtained with sera from uninfected individuals is considered positive.

          Antigenic portions of *M. tuberculosis* antigens may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited  
15       therein. Such techniques include screening polypeptide portions of the native antigen for antigenic properties. The representative ELISAs described herein may generally be employed in these screens. An antigenic portion of a polypeptide is a portion that, within such representative assays, generates a signal in such assays that is substantially similar to that generated by the full length antigen. In other words, an antigenic portion  
20       of a *M. tuberculosis* antigen generates at least about 20%, and preferably about 100%, of the signal induced by the full length antigen in a model ELISA as described herein.

          Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated  
25       using techniques well known in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963).

according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

5           Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available  
10   filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

          Any of a variety of expression vectors known to those of ordinary skill in  
15   the art may be employed to express recombinant polypeptides as described herein. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian  
20   cell line, such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

          In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the polypeptides  
25   are at least about 30% pure, more preferably at least about 90% pure and most preferably at least about 99% pure. For use in the methods described herein, however, such substantially pure polypeptides may be combined

          For a description of the methods of preparation of the polypeptides, see the following:

antigen (or a variant of such an antigen), where the antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);
- 5 (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);
- 10 (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);
- 15 (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123);
- 20 (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser (SEQ ID NO: 124);
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp (SEQ ID NO: 130) or
- 25 (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly (SEQ ID NO: 131)

wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence

encoding the antigen identified as (a) above is provided in SEQ ID NO: 96; its deduced amino acid sequence is provided in SEQ ID NO: 97. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID NO: 24, a DNA sequence corresponding to antigen (c) is provided in SEQ ID NO: 25 and a DNA sequence corresponding to  
5 antigen (f) is disclosed in SEQ ID NO: 94 and its deduced amino acid sequence is provided in SEQ ID NO: 95.

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only  
10 in conservative substitutions and/or modifications:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132) or

(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

15 wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence encoding the antigen of (n) above is provided in SEQ ID NO: 235, with the corresponding predicted full-length amino acid sequence being provided in SEQ ID NO: 236.

20 In other specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID NOS 1, 2, 4, 10, 13-25, 52, 94 and 96, (b) the complements of such DNA sequences, or (c) DNA sequences  
25 substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen) which may or may not be soluble that comprises one or

242, 248-251, 256-271, 287, 288, 290-293 and 298-337, (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In a related aspect, the present invention provides fusion proteins  
5 comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen described in Andersen and Hansen, *Infect Immun.* 57:2481-2488, 1989, (Genbank Accession No. M30046) or ESAT-6 (SEQ ID NOS: 98 and 99), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a  
10 linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or  
15 without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the  
20 second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and  
25 (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids such as Thr, Ser, Met, Ala, Val, Leu, Ile, and Phe

*Natl. Acad. Sci. USA* 83:8258-8562, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric hindrance.

In another aspect, the present invention provides methods for using the polypeptides described above to diagnose tuberculosis. In this aspect, methods are provided for detecting *M. tuberculosis* infection in a biological sample, using one or more of the above polypeptides, alone or in combination. In embodiments in which multiple polypeptides are employed, polypeptides other than those specifically described herein, such as the 38 kD antigen described in Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989, may be included. As used herein, a "biological sample" is any antibody-containing sample obtained from a patient. Preferably, the sample is whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid or urine. More preferably, the sample is a blood, serum or plasma sample obtained from a patient or a blood supply. The polypeptide(s) are used in an assay, as described below, to determine the presence or absence of antibodies to the polypeptide(s) in the sample, relative to a predetermined cut-off value. The presence of such antibodies indicates previous sensitization to mycobacterial antigens which may be indicative of tuberculosis.

In embodiments in which more than one polypeptide is employed, the polypeptides used are preferably complementary (i.e., one component polypeptide will tend to detect infection in samples where the infection would not be detected by another component polypeptide). Complementary polypeptides may generally be identified by using each polypeptide individually to evaluate serum samples obtained from a series of patients known to be infected with *M. tuberculosis*. After determining which samples test positive (as described below) with each polypeptide, combinations of two or more polypeptides may be formulated that are capable of detecting infection in most, or all, of the samples tested. Such polypeptides are complementary. For example, approximately

polypeptides may, therefore, be used in combination with the 38 kD antigen to improve sensitivity of a diagnostic test.

There are a variety of assay formats known to those of ordinary skill in the art for using one or more polypeptides to detect antibodies in a sample. See, e.g.,  
5 Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, which is incorporated herein by reference. In a preferred embodiment, the assay involves the use of polypeptide immobilized on a solid support to bind to and remove the antibody from the sample. The bound antibody may then be detected using a detection reagent that contains a reporter group. Suitable detection reagents include  
10 antibodies that bind to the antibody/polypeptide complex and free polypeptide labeled with a reporter group (e.g., in a semi-competitive assay). Alternatively, a competitive assay may be utilized, in which an antibody that binds to the polypeptide is labeled with a reporter group and allowed to bind to the immobilized antigen after incubation of the antigen with the sample. The extent to which components of the sample inhibit the  
15 binding of the labeled antibody to the polypeptide is indicative of the reactivity of the sample with the immobilized polypeptide.

The solid support may be any solid material known to those of ordinary skill in the art to which the antigen may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane.  
20 Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The polypeptides may be bound to the solid support using a variety of  
25 techniques known to those of ordinary skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "bound" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the antigen and functional groups on the

be achieved by contacting the polypeptide, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of polypeptide ranging  
5 from about 10 ng to about 1  $\mu$ g, and preferably about 100 ng, is sufficient to bind an adequate amount of antigen.

Covalent attachment of polypeptide to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the  
10 polypeptide. For example, the polypeptide may be bound to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the polypeptide (see, e.g., Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is an enzyme linked immunosorbent  
15 assay (ELISA). This assay may be performed by first contacting a polypeptide antigen that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that antibodies to the polypeptide within the sample are allowed to bind to the immobilized polypeptide. Unbound sample is then removed from the immobilized polypeptide and a detection reagent capable of binding to the immobilized  
20 antibody-polypeptide complex is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific detection reagent.

More specifically, once the polypeptide is immobilized on the support as described above, the remaining protein binding sites on the support are typically  
25 blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO) may be employed. The immobilized polypeptide is then incubated with the sample, and antibody is allowed to bind to the antigen. The sample may be diluted with a suitable

detect the presence of antibody within a *M. tuberculosis*-infected sample. Preferably, the contact time is sufficient to achieve a level of binding that is at least 95% of that achieved at equilibrium between bound and unbound antibody. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily  
5 determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. Detection reagent may then be added to the solid support. An appropriate detection reagent is any  
10 compound that binds to the immobilized antibody-polypeptide complex and that can be detected by any of a variety of means known to those in the art. Preferably, the detection reagent contains a binding agent (such as, for example, Protein A, Protein G, immunoglobulin, lectin or free antigen) conjugated to a reporter group. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors,  
15 inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups, biotin and colloidal particles, such as colloidal gold and selenium. The conjugation of binding agent to reporter group may be achieved using standard methods known to those of ordinary skill in the art. Common binding agents may also be purchased conjugated to a variety of reporter groups from many commercial sources (e.g., Zymed Laboratories,  
20 San Francisco, CA, and Pierce, Rockford, IL).

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound antibody. An appropriate amount of time may generally be determined from the manufacturer's instructions or by assaying the level of binding that occurs over a period of time.  
25 Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic

radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of anti-*M. tuberculosis* antibodies  
5 in the sample, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antigen is incubated with samples from an uninfected patient. In general, a sample generating a signal that is three standard deviations above  
10 the predetermined cut-off value is considered positive for tuberculosis. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, pp. 106-107. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive  
15 rates (i.e., sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (i.e., the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the  
20 cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for tuberculosis.

In a related embodiment, the assay is performed in a rapid flow-through  
25 or strip test format, wherein the antigen is immobilized on a membrane, such as nitrocellulose. In the flow-through test, antibodies within the sample bind to the immobilized polypeptide as the sample passes through the membrane. A detection reagent (e.g., protein A colloidal gold) then binds to the antibody-polypeptide complex.

strip test format, one end of the membrane to which polypeptide is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing detection reagent and to the area of immobilized polypeptide. Concentration of detection reagent at the polypeptide indicates the presence of anti-  
5 *M. tuberculosis* antibodies in the sample. Typically, the concentration of detection reagent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of polypeptide immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of antibodies that would be  
10 sufficient to generate a positive signal in an ELISA, as discussed above. Preferably, the amount of polypeptide immobilized on the membrane ranges from about 25 ng to about 1 µg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount (e.g., one drop) of patient serum or blood.

Of course, numerous other assay protocols exist that are suitable for use  
15 with the polypeptides of the present invention. The above descriptions are intended to be exemplary only.

In yet another aspect, the present invention provides antibodies to the inventive polypeptides. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. See, e.g., Harlow and Lane, *Antibodies: A  
20 Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one such technique, an immunogen comprising the antigenic polypeptide is initially injected into any of a wide variety of mammals (e.g., mice, rats, rabbits, sheep and goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response  
25 may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies

Monoclonal antibodies specific for the antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process, for example, an affinity chromatography step.

Antibodies may be used in diagnostic tests to detect the presence of *M. tuberculosis* antigens using assays similar to those detailed above and other techniques well known to those of skill in the art, thereby providing a method for detecting *M. tuberculosis* infection in a patient.

thereof. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify *M. tuberculosis*-specific cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for a DNA molecule encoding a polypeptide of the present invention.

5 The presence of the amplified cDNA is then detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes specific for a DNA molecule encoding a polypeptide of the present invention may be used in a hybridization assay to detect the presence of an inventive polypeptide in a biological sample.

10 As used herein, the term "oligonucleotide primer/probe specific for a DNA molecule" means an oligonucleotide sequence that has at least about 80%, preferably at least about 90% and more preferably at least about 95%, identity to the DNA molecule in question. Oligonucleotide primers and/or probes which may be usefully employed in the inventive diagnostic methods preferably have at least about  
15 10-40 nucleotides. In a preferred embodiment, the oligonucleotide primers comprise at least about 10 contiguous nucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Preferably, oligonucleotide probes for use in the inventive diagnostic methods comprise at least about 15 contiguous oligonucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Techniques for  
20 both PCR based assays and hybridization assays are well known in the art (see, for example, Mullis *et al.* *Ibid*; Ehrlich, *Ibid*). Primers or probes may thus be used to detect *M. tuberculosis*-specific sequences in biological samples. DNA probes or primers comprising oligonucleotide sequences described above may be used alone, in combination with each other, or with previously identified sequences, such as the 38 kD antigen discussed above.

The following examples are offered by way of illustration and not by way of limitation.

## EXAMPLES

### EXAMPLE 1

#### PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES

#### FROM *M. TUBERCULOSIS* CULTURE FILTRATE

This example illustrates the preparation of *M. tuberculosis* soluble polypeptides from culture filtrate. Unless otherwise noted, all percentages in the following example are weight per volume.

*M. tuberculosis* (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45  $\mu$  filter into a sterile 2.5 L bottle. The media was then filtered through a 0.2  $\mu$  filter into a sterile 4 L bottle.  $\text{NaN}_3$  was then added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

The culture filtrate was concentrated by placing the filtrate in a 12 L reservoir that had been autoclaved and feeding the filtrate into a 400 ml Amicon stir cell which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was then dialyzed into 0.1% ammonium bicarbonate using a 5,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were then dialyzed against 0.01 mM 1,3-bis(tris(hydroxymethyl) methylamino)propane, pH 7.5 (Bis-Tris propane buffer), the initial conditions for ion exchange chromatography. Fractionation was performed using

Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were  
5. dialyzed against distilled water and lyophilized. The resulting material was dissolved in 0.1% trifluoroacetic acid (TFA) pH 1.9 in water, and the polypeptides were purified on a Delta-Pak C18 column (Waters, Milford, MA) 300 Angstrom pore size, 5 micron particle size (3.9 x 150 mm). The polypeptides were eluted from the column with a linear gradient from 0-60% dilution buffer (0.1% TFA in acetonitrile). The flow rate  
10 was 0.75 ml/minute and the HPLC eluent was monitored at 214 nm. Fractions containing the eluted polypeptides were collected to maximize the purity of the individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD  
15 skin test positive and whose T cells were shown to proliferate in response to PPD and crude soluble proteins from MTB were cultured in medium comprising RPMI 1640 supplemented with 10% pooled human serum and 50 µg/ml gentamicin. Purified polypeptides were added in duplicate at concentrations of 0.5 to 10 µg/mL. After six days of culture in 96-well round-bottom plates in a volume of 200 µl, 50 µl of medium  
20 was removed from each well for determination of IFN-γ levels, as described below. The plates were then pulsed with 1 µCi/well of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

25 IFN-γ was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to human IFN-γ (Chemicon) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 8% (W/V) non-fat dried milk for 1 hour at room

room temperature. The plates were again washed and a polyclonal rabbit anti-human IFN- $\gamma$  serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Jackson Labs.) was added at a 1:2000  
5 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added. The reaction was stopped after 20 min with 1 N sulfuric acid. Optical density was determined at 450 nm using 570 nm as a reference wavelength. Fractions that resulted in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone,  
10 plus 3 standard deviations, were considered positive.

For sequencing, the polypeptides were individually dried onto Biobrene™ (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced  
15 from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide by comparing the retention time of the PTH amino acid derivative to the appropriate PTH derivative standards.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

- 20 (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Ihr-Xaa-Asn-Iyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 54);
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 55);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 56);  
25
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 57);
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val

- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 59);
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala (SEQ ID NO: 60); and
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 61);

wherein Xaa may be any amino acid.

An additional antigen was isolated employing a microbore HPLC purification step in addition to the procedure described above. Specifically, 20  $\mu$ l of a fraction comprising a mixture of antigens from the chromatographic purification step previously described, was purified on an Aquapore C18 column (Perkin Elmer/Applied Biosystems Division, Foster City, CA) with a 7 micron pore size, column size 1 mm x 100 mm, in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1%/minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80  $\mu$ l/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller components and a polypeptide was obtained which was shown to have a molecular weight of 12.054 Kd (by mass spectrometry) and the following N-terminal sequence.

- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-Val-Asp (SEQ ID NO: 62)

This polypeptide was shown to induce proliferation and IFN- $\gamma$  production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following analysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using ion exchange chromatography on a Poros Q1 column 4.6 x 100 mm.

were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and resuspended in 80  $\mu$ l of aqueous 0.1% TFA and further subjected to reverse phase chromatography on a Vydac C4 column 4.6 x 150 mm (Western Analytical, Temecula, CA) with a linear gradient of 0-100% acetonitrile (0.1% TFA) at a flow rate of 2 ml/min. Eluent was monitored at 214 nm.

The fraction with biological activity was separated into one major peak plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequences, respectively:

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) and

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131), wherein Xaa may be any amino acid

Using the assays described above, these polypeptides were shown to induce proliferation and IFN- $\gamma$  production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a), (c), (d) and (g) above were obtained by screening a *M. tuberculosis* genomic library using <sup>32</sup>P end

corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID NO: 96. The polypeptide encoded by SEQ ID NO: 96 is provided in SEQ ID NO: 97. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID NO: 52. The polypeptide encoded by SEQ ID NO: 52 is provided in SEQ ID NO: 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID NO: 24, and the screen performed with a probe corresponding to antigen (c) identified a clone having the sequence provided in SEQ ID NO: 25.

The above amino acid sequences were compared to known amino acid sequences in the gene bank using the DNA STAR system. The database searched contains some 173,000 proteins and is a combination of the Swiss, PIR databases along with translated protein sequences (Version 87). No significant homologies to the amino acid sequences for antigens (a)-(h) and (l) were detected.

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen an *M. tuberculosis* library and a full length copy of the *M. tuberculosis* homologue was obtained (SEQ ID NO: 94).

The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN- $\gamma$  assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1.

**TABLE 1**  
**RESULTS OF PBMC PROLIFERATION AND IFN- $\gamma$  ASSAYS**

Sequence	Proliferation	IFN- $\gamma$
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4 (compared to cells cultured in medium alone) were scored as +, as SI of 4-8 or 2-4 at a concentration of 1  $\mu$ g or less was scored as ++ and an SI of greater than 8 was scored as +++. The antigen of sequence (i) was found to have a high SI (++++) for one donor and lower SI (— and —) for the two other donors in both proliferation and IFN- $\gamma$  assays. These results indicate that these antigens are capable of inducing proliferation and/or interferon- $\gamma$  production.

#### EXAMPLE 2 USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.

Desiccated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with

DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with  $\alpha$ -D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any amino acid.

Comparison of this sequence with those in the gene bank as described above, revealed no significant homologies to known sequences.

A DNA sequence that encodes the antigen designated as (m) above was obtained by screening a genomic *M. tuberculosis* Erdman strain library using labeled degenerate oligonucleotides corresponding to the N-terminal sequence of SEQ ID NO:137. A clone was identified having the DNA sequence provided in SEQ ID NO: 198. This sequence was found to encode the amino acid sequence provided in SEQ ID NO: 199. Comparison of these sequences with those in the genebank revealed some similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

#### EXAMPLE 3

##### PREPARATION OF DNA SEQUENCES ENCODING *M. tuberculosis* ANTIGENS

This example illustrates the preparation of DNA sequences encoding *M. tuberculosis* antigens by screening a *M. tuberculosis* expression library with sera obtained from patients infected with *M. tuberculosis*.

A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* SUPERNATANT

Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The DNA was randomly sheared and used to construct an expression library using the Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis* cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of protein antigen in a total volume of 2 ml containing 100 µg muramyl dipeptide (Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's adjuvant. Finally, the rabbit was immunized intravenously four weeks later with 50 µg protein antigen. The anti-sera were used to screen the expression library as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty two clones were purified. Of these, 25 represent sequences that have not been previously identified in *M. tuberculosis*. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med.* 187:1527-1537, 1995. Representative partial sequences of DNA molecules identified in this screen are provided in SEQ ID NOS 1-25. The corresponding predicted amino acid sequences are shown in SEQ ID NOS 64-88.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID NOS 77, 69, 71, 70) show some homology to sequences previously identified in *Mycobacterium leprae* but not in *M. tuberculosis*. TbRA2A (SEQ ID NOS 77) is identical to

previously identified in *M. tuberculosis*. No significant homologies were found to TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRA19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID NOS: 64, 78, 82, 83, 65, 68, 76, 72, 76, 79, 81, 80, 67, respectively). The clone  
5 TbRa24 is overlapping with clone TbRa29

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial Sau3A digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with  
10 active pulmonary or pleural disease, were used in the expression screening. The pools were designated TbL, TbM and TbH, referring to relative reactivity with H37Ra lysate (i.e., TbL = low reactivity, TbM = medium reactivity and TbH = high reactivity) in both ELISA and immunoblot format. A fourth pool of sera from seven patients with active  
15 pulmonary tuberculosis was also employed. All of the sera lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified.  
20 Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty-two clones were purified. Of these, 11 represented sequences that had not been previously identified in human *M. tuberculosis*.

NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID NOS. 89-93. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infect. Immun.* 63:1710-1717, 1995).

Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS. 107, 108, 111, 113, and 114). (SEQ ID NOS. 107 and 108 are non-contiguous sequences from clone Tb38-1F2.) Two open reading frames were deduced in Tb38-1F2: one corresponds to Tb37FL (SEQ. ID. NO. 109), the second, a partial sequence, may be the homologue of Tb38-1 and is called Tb38-IN (SEQ. ID NO. 110). The deduced amino acid sequence of Tb38-1F3 is presented in SEQ. ID. NO. 112. A TbH-9 probe identified three clones in the H37Rv library: TbH-9-FL (SEQ. ID NO. 101), which may be the homologue of TbH-9 (H37Ra), TbH-9-1 (SEQ. ID NO. 103), and TbH-8-2 (SEQ. ID NO. 105) is a partial clone of TbH-8. The deduced amino acid sequences for these three clones are presented in SEQ. ID NOS. 102, 104 and 106.

Further screening of the *M. tuberculosis* genomic DNA library, as described above, resulted in the recovery of ten additional reactive clones, representing seven different genes. One of these genes was identified as the 38 Kd antigen discussed above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined

TbH-33) are provided in SEQ ID NO: 133-136, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 137-140, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29 (which contains the reactive open reading frame), although the 3' end of TbH-29 was found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

Positive phagemid from this additional screening were used to infect *E. coli* XL-1 Blue MRF', as described in Sambrook et al., *supra*. Induction of recombinant protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human *M. tuberculosis* sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of <sup>125</sup>I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours to 11 days. The results of the immunoblots are summarized in Table 2.

TABLE 2

Antigen	Human <i>M. tb</i> Sera	Anti-lacZ Sera
TbH-29	45 Kd	45 Kd
TbH-30	No reactivity	29 Kd
TbH-32	12 Kd	12 Kd
TbH-33	16 Kd	16 Kd

the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera (panel I) and antisera (panel II) against secretory proteins, recombinant 85b, recombinant Tb38-1 and recombinant TbH-9 are shown in Figures 2A-D, respectively, wherein the lane designations are as follows: 1) molecular weight protein standards; 2) 5 µg of *M. tuberculosis* lysate; 3) 5 µg secretory proteins; 4) 50 ng recombinant Tb38-1; 5) 50 ng recombinant TbH-9; and 6) 50 ng recombinant 85b. The recombinant antigens were engineered with six terminal histidine residues and would therefore be expected to migrate with a mobility approximately 1 kD larger than the native protein. In Figure 2B, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T-cell clones with *M. tuberculosis* lysate.

The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen, TbRa11, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 3A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of *M. tuberculosis* secretory proteins. Figure 3B shows the production of IFN- $\gamma$  by a second TbH-9-specific T cell clone (designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by *M. tuberculosis*.

C. USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis, as described above in Example 3B, with the secondary antibody being goat anti-human IgG - A - M (H+L) conjugated with alkaline phosphatase.

Eighteen clones were purified. Of these, 4 clones (hereinafter referred to as XP14, XP24, XP31 and XP32) were found to bear some similarity to known sequences. The determined DNA sequences for XP14, XP24 and XP31 are provided in SEQ ID NOS. 151-153, respectively, with the 5' and 3' DNA sequences for XP22 being provided in SEQ ID NOS. 154 and 155, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID NO: 156. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID NO: 157.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6, XP17-XP19, XP22, XP25, XP27, XP30 and XP36) with those

NOS: 158 and 159, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID NOS: 160-163, respectively, and the 5' and 3' sequences for XP2, XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID NOS: 164 and 165; 166 and 167; 168 and 169; 170 and 171; 172 and 173; 174 and 175; and 176 and 177, respectively. XP1 was found to overlap with the DNA sequences for TbH4, disclosed above. The full-length DNA sequence for TbH4-XP1 is provided in SEQ ID NO: 178. This DNA sequence was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 179. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 180. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID NOS: 181 and 182, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 183.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. Recombinant XP1 was found to stimulate cell proliferation and IFN- $\gamma$  production in T cells isolated from an *M. tuberculosis*-immune donors.

#### D. USE OF LYSATE POSITIVE SERUM POOL FROM PATIENTS HAVING TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda screen expression system (Novagen, Madison, WI), as described below in Example 4. Pooled serum obtained from *M. tuberculosis*-infected patients and that was shown to react with *M. tuberculosis* lysate but not with the previously expressed proteins 38kD, Tb38-1, TbRa3, TbH4, DPEP and TbRa11, was used to screen the expression library as described above in Example 3B, with the secondary antibody being anti-mouse IgG.

(hereinafter referred to as LSER-10, LSER-11, LSER-12, LSER-13, LSER-16, LSER-18, LSER-23, LSER-24, LSER-25 and LSER-27). The determined 5' cDNA sequences for LSER-10, LSER-11, LSER-12, LSER-13, LSER-16 and LSER-25 are provided in SEQ ID NO: 237-242, respectively, with the corresponding predicted amino acid sequences for LSER-10, LSER-12, LSER-13, LSER-16 and LSER-25 being provided in SEQ ID NO: 243-247, respectively. The determined full-length cDNA sequences for LSER-18, LSER-23, LSER-24 and LSER-27 are shown in SEQ ID NO: 248-251, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 252-255. The remaining seventeen clones were found to show similarities to unknown sequences previously identified in *M. tuberculosis*. The determined 5' cDNA sequences for sixteen of these clones (hereinafter referred to as LSER-1, LSER-3, LSER-4, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30) are provided in SEQ ID NO: 256-271, respectively, with the corresponding predicted amino acid sequences for LSER-1, LSER-3, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30 being provided in SEQ ID NO: 272-286, respectively. The determined full-length cDNA sequence for the clone LSER-9 is provided in SEQ ID NO: 287. The reverse complement of LSER-9 (SEQ ID NO: 288) was found to encode the predicted amino acid sequence of SEQ ID NO: 289.

#### 5. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTISERA RAISED AGAINST *M. TUBERCULOSIS* FRACTIONATED PROTEINS

*M. tuberculosis* lysate was prepared as described above in Example 2. The resulting material was fractionated by HPLC and the fractions screened by Western blot for serological activity with a serum pool from *M. tuberculosis*-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-*M. tuberculosis* serum was prepared by immunizing rabbits with the

Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen, HSP60. Of the remaining eight clones, six (hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF12) were found to bear some similarity to previously identified *M. tuberculosis* sequences. The determined DNA sequences for RDIF2, RDIF5, RDIF8, RDIF10 and RDIF11 are provided in SEQ ID NOS. 184-188, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NOS. 189-193, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID NOS. 194 and 195, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID NOS. 196 and 197, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6, RDIF8, RDIF10 and RDIF11 were prepared as described above. These antigens were found to stimulate cell proliferation and IFN- $\gamma$  production in T cells isolated from *M. tuberculosis*-immune donors.

#### EXAMPLE 1

##### PURIFICATION AND CHARACTERIZATION OF A POLYPEPTIDE FROM TUBERCULIN PURIFIED PROTEIN DERIVATIVE

An *M. tuberculosis* polypeptide was isolated from tuberculin purified protein derivative (PPD) as follows:

PPD was prepared as published with some modification (Seibert, E. et

Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100°C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22  $\mu$  filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP-HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perceptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 10 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80  $\mu$ l/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the amino terminal using a Perkin Elmer Applied Biosystems Division Procise 492 protein sequencer as described above and found to have the N terminal sequence shown in SEQ ID NO: 124. Comparison of this sequence with known sequences in the gene bank as described above revealed no known homologues. Four cyanogen bromide fragments of DPPD were isolated and found to have the sequences shown in SEQ ID

sequence with a sequence present within the *M. tuberculosis* cosmid MTY21C12. An open reading frame of 336 bp was identified. The full-length DNA sequence for DPPD is provided in SEQ ID NO: 235, with the corresponding full-length amino acid sequence being provided in SEQ ID NO: 236.

5

### EXAMPLE 5

#### USE OF SERA FROM TUBERCULOSIS-INFECTED MONKEYS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

10

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). Serum samples were obtained from a cynomolgous monkey 18, 33, 51 and 56 days following infection with *M. tuberculosis* Erdman strain. These samples were pooled and used to screen the *M. tuberculosis* genomic DNA expression library using the procedure described above in Example 3C.

Twenty clones were purified. The determined 5' DNA sequences for the clones referred to as MO-1, MO-2, MO-4, MO-8, MO-9, MO-26, MO-28, MO-29, MO-30, MO-34 and MO-35 are provided in SEQ ID NO: 210-220, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 221-231. The full-length DNA sequence of the clone MO-10 is provided in SEQ ID NO: 232, with the corresponding predicted amino acid sequence being provided in SEQ ID NO: 233. The 3' DNA sequence for the clone MO-27 is provided in SEQ ID NO: 234.

Clones MO-1, MO-30 and MO-35 were found to show a high degree of relatedness and showed some homology to a previously identified unknown *M. tuberculosis* sequence and to cosmid MTC1237. MO-2 was found to show some homology to aspartokinase from *M. tuberculosis*. Clones MO-3, MO-7 and MO-27 were found to be identical and to show a high degree of relatedness to MO-5. All clones

and MO-34 were found to show some homology to cosmid SCY21B4 and *M. smegmatis* integration host factor, and were both found to show some homology to a previously identified, unknown *M. tuberculosis* sequence. MO-6 was found to show some homology to *M. tuberculosis* heat shock protein 65. MO-8, MO-9, MO-10, MO-26 and MO-29 were found to be highly related to each other and to show some homology to *M. tuberculosis* dihydrolipamide succinyltransferase. MO-28, MO-31 and MO-32 were found to be identical and to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a previously identified 14 kDa *M. tuberculosis* heat shock protein

Further studies using the above protocol resulted in the isolation of an additional four clones, hereinafter referred to as MO-12, MO-13, MO-19 and MO-39. The determined 5' cDNA sequences for these clones are provided in SEQ ID NO: 290-293, respectively, with the corresponding predicted protein sequences being provided in SEQ ID NO: 294-297, respectively. Comparison of these sequences with those in the gene bank as described above revealed no significant homologies to MO-39. MO-12, MO-13 and MO-19 were found to show some homologies to unknown sequences previously isolated from *M. tuberculosis*.

#### EXAMPLE 9

##### ISOLATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS BY SCREENING OF A NOVEL EXPRESSION LIBRARY

This example illustrates isolation of DNA sequences encoding *M. tuberculosis* antigens by screening of a novel expression library with sera from *M. tuberculosis*-infected patients that were shown to be unreactive with a panel of the recombinant *M. tuberculosis* antigens TbRa11, TbRa3, Tb38-1, TbH4, TbF and 38 kD

Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase (Boehringer

extract (Novagen). The resulting library was screened with sera from several *M. tuberculosis* donors that had been shown to be negative on a panel of previously identified *M. tuberculosis* antigens as described above in Example 3B.

A total of 22 different clones were isolated. By comparison, screening of the  $\lambda$ Zap library described above using the same sera did not result in any positive hits. One of the clones was found to represent TbRa11, described above. The determined 5' cDNA sequences for 19 of the remaining 21 clones (hereinafter referred to as Erdsn1, Erdsn2, Erdsn4-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25) are provided in SEQ ID NO: 298-317, respectively, with the determined 3' cDNA sequences for Erdsn1, Erdsn2, Erdsn4, Erdsn5, Erdsn7-Erdsn10, Erdsn12-Erdsn18, Erdsn21-Erdsn23 and Erdsn25 being provided in SEQ ID NO: 318-336, respectively. The complete cDNA insert sequence for the clone Erdsn24 is provided in SEQ ID NO: 337. Comparison of the determined cDNA sequences with those in the gene bank revealed no significant homologies to the sequences provided in SEQ ID NO: 304, 311, 313-315, 317, 319, 324, 326, 329, 331, 333, 335 and 337. The sequences of SEQ ID NO: 298-303, 305-310, 312, 316, 318, 320-321, 324-326, 328, 330, 332, 334 and 336 were found to show some homology to unknown sequences previously identified in *M. tuberculosis*.

#### EXAMPLE 7

##### ISOLATION OF SOLUBLE *M. TUBERCULOSIS* ANTIGENS

##### USING MASS SPECTROMETRY

This example illustrates the use of mass spectrometry to identify soluble *M. tuberculosis* antigens.

In a first approach, *M. tuberculosis* culture filtrate was screened by Western analysis using serum from a tuberculosis-infected individual. The reactive bands were excised from a silver stained gel and the amino acid sequences determined

the gene bank revealed homology to the 85b precursor antigen previously identified in *M. tuberculosis*.

In a second approach, the high molecular weight region of *M. tuberculosis* culture supernatant was studied. This area may contain immunodominant antigens which may be useful in the diagnosis of *M. tuberculosis* infection. Two known monoclonal antibodies, IT42 and IT57 (available from the Center for Disease Control, Atlanta, GA), show reactivity by Western analysis to antigens in this vicinity, although the identity of the antigens remains unknown. In addition, unknown high-molecular weight proteins have been described as containing a surrogate marker for *M. tuberculosis* infection in HIV-positive individuals (*Jnl. Infect. Dis.*, 176:133-143, 1997). To determine the identity of these antigens, two-dimensional gel electrophoresis and two-dimensional Western analysis were performed using the antibodies IT57 and IT42. Five protein spots in the high molecular weight region were identified, individually excised, enzymatically digested and subjected to mass spectrometric analysis.

The determined amino acid sequences for three of these spots (referred to as spots 1, 2 and 4) are provided in SEQ ID NO: 339, 340-341 and 342, respectively. Comparison of these sequences with those in the gene bank revealed that spot 1 is the previously identified *PcK-1*, a phosphoenolpyruvate kinase. The two sequences isolated from spot 2 were determined to be from two DNAs, previously identified in *M. tuberculosis* as heat shock proteins. Spot 4 was determined to be the previously identified *M. tuberculosis* protein *Kat G*. To the best of the inventors' knowledge, neither *PcK-1* nor the two DNAs have previously been shown to have utility in the diagnosis of *M. tuberculosis* infection.

### EXAMPLE 5

#### SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide

attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-*t*-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

This procedure was used to synthesize a TbM-1 peptide that contains one and a half repeats of a TbM-1 sequence. The TbM-1 peptide has the sequence GCGDRSGGNLDQIRLRRDRSGGNL (SEQ ID NO: 63).

#### EXAMPLE 9

##### USE OF REPRESENTATIVE ANTIGENS FOR SERODIAGNOSIS OF TUBERCULOSIS

This Example illustrates the diagnostic properties of several representative antigens.

Assays were performed in 96-well plates were coated with 200 ng antigen diluted to 50  $\mu$ L in carbonate coating buffer, pH 9.6. The wells were coated overnight at 4°C or 2 hours at 37°C. The plate contents were then removed and the wells were blocked for 2 hours with 200  $\mu$ L of PBS/1% BSA. After the blocking step, the wells were washed five times with PBS/0.1% Tween 20™. 50  $\mu$ L sera, diluted 1:100 in PBS/0.1% Tween 20™/0.1% BSA, was then added to each well and incubated for 30 minutes at room temperature. The plates were then washed again five times with PBS/0.1% Tween 20™.

μL of the diluted conjugate was added to each well and incubated for 30 minutes at room temperature. Following incubation, the wells were washed five times with PBS/0.1% Tween 20™. 100 μL of tetramethylbenzidine peroxidase (TMB) substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD) was added, undiluted, and  
5 incubated for about 15 minutes. The reaction was stopped with the addition of 100 μL of 1 N H<sub>2</sub>SO<sub>4</sub> to each well, and the plates were read at 450 nm.

Figure 4 shows the ELISA reactivity of two recombinant antigens isolated using method A in Example 3 (TbRa3 and TbRa9) with sera from *M. tuberculosis* positive and negative patients. The reactivity of these antigens is  
10 compared to that of bacterial lysate isolated from *M. tuberculosis* strain H37Ra (Difco, Detroit, MI). In both cases, the recombinant antigens differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 56 out of 87 positive sera, and TbRa9 detected 111 out of 165 positive sera.

Figure 5 illustrates the ELISA reactivity of representative antigens  
15 isolated using method B of Example 3. The reactivity of the recombinant antigens TbH4, TbH12, Tb38-1 and the peptide TbM-1 (as described in Example 4) is compared to that of the 38 kD antigen described by Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989. Again, all of the polypeptides tested differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbH4  
20 detected 97 out of 126 positive sera, TbH12 detected 80 out of 125 positive sera, 38-1 detected 91 out of 101 positive sera and the TbM-1 peptide detected 25 out of 30 positive sera.

The reactivity of four antigens (TbRa1, TbRa9, TbH4 and TbH12) with sera from a group of *M. tuberculosis* infected patients with differing reactivity in the  
25 acid fast stain of sputum (Smithwick and David, *Tubercle* 52:226, 1971) was also examined, and compared to the reactivity of *M. tuberculosis* lysate and the 38 kD antigen. The results are presented in Table 3, below.

TABLE 3  
 REACTIVITY OF ANTIGENS WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-2	----	1.853	0.634	0.998	1.022	1.030	1.314
Tb01B93I-19	-----	2.657	2.322	0.608	0.837	1.857	2.335
Tb01B93I-8	---	2.703	0.527	0.492	0.281	0.501	2.002
Tb01B93I-19	---	1.565	1.301	0.685	0.216	0.448	0.458
Tb01B93I-11	---	2.817	0.697	0.509	0.301	0.173	2.608
Tb01B93I-15	---	1.28	0.283	0.808	0.218	1.537	0.811
Tb01B93I-16	---	2.908	>3	0.899	0.441	0.593	1.080
Tb01B93I-25	---	0.395	0.131	0.335	0.211	0.107	0.948
Tb01B93I-87	---	2.653	2.432	2.282	0.977	1.221	0.857
Tb01B93I-89	---	1.912	2.370	2.436	0.876	0.520	0.952
Tb01B94I-108	---	1.639	0.341	0.797	0.368	0.654	1.798
Tb01B94I-201	---	1.721	0.410	0.661	0.137	0.064	0.692
Tb01B93I-88	--	1.930	1.269	2.519	1.381	0.214	0.530
Tb01B93I-92	--	2.155	2.329	2.78	0.685	0.997	2.527
Tb01B94I-109	-	0.993	0.620	0.574	0.441	0.5	2.558
Tb01B94I-210	---	2.777	>3	0.393	0.367	1.004	1.315
Tb01B94I-224	---	2.913	0.476	0.251	1.297	1.990	0.256
Tb01B93I-9	-	2.649	0.278	0.210	0.140	0.181	1.586

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-22	+	0.714	0.451	2.082	0.285	0.269	1.159
Tb01B93I-31	+	0.956	0.490	1.019	0.812	0.176	1.293
Tb01B93I-32	-	2.261	0.786	0.668	0.273	0.535	0.405
Tb01B93I-52		0.658	0.114	0.434	0.330	0.273	1.140
Tb01B93I-99	-	2.118	0.584	1.62	0.119	0.977	0.729
Tb01B94I-130		1.349	0.224	0.86	0.282	0.583	2.146
Tb01B94I-131	-	0.685	0.324	1.173	0.059	0.118	1.431
AT4-0070	Normal	0.072	0.043	0.092	0.071	0.040	0.039
AT4-0105	Normal	0.397	0.121	0.118	0.103	0.078	0.390
3-15/94-1	Normal	0.227	0.064	0.098	0.026	0.001	0.228
4-15/93-2	Normal	0.114	0.240	0.071	0.034	0.041	0.264
5-26/94-4	Normal	0.089	0.259	0.096	0.046	0.008	0.053
5-26/94-3	Normal	0.139	0.093	0.085	0.019	0.067	0.01

Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 23 out of 27 positive sera, TbRa9 detected 22 out of 27, TbH4 detected 18 out of 27 and TbH12 detected 15 out of 27. If used in combination, these four antigens would have a theoretical sensitivity of 27 out of 27, indicating that these antigens should complement each other in the serological detection of *M. tuberculosis* infection. In addition, several of the recombinant antigens detected positive sera that were not detected using the 38 kD antigen, indicating that these antigens may be complementary to the 38 kD antigen.

The results are shown in Figure 6 which indicates that TbRa11, while being negative with sera from PPD positive and normal donors, detected sera that were negative with the 38 kD antigen. Of the thirteen 38 kD negative sera tested, nine were positive with TbRa11, indicating that this antigen may be reacting with a sub-group of 38 kD antigen negative sera. In contrast, in a group of 38 kD positive sera where TbRa11 was reactive, the mean OD 450 for TbRa11 was lower than that for the 38 kD antigen. The data indicate an inverse relationship between the presence of TbRa11 activity and 38 kD positivity.

The antigen TbRa2A was tested in an indirect ELISA using initially 50  $\mu$ l of serum at 1:100 dilution for 30 minutes at room temperature followed by washing in PBS Tween and incubating for 30 minutes with biotinylated Protein A (Zymed, San Francisco, CA) at a 1:10,000 dilution. Following washing, 50  $\mu$ l of streptavidin-horseradish peroxidase (Zymed) at 1:10,000 dilution was added and the mixture incubated for 30 minutes. After washing, the assay was developed with TMB substrate as described above. The reactivity of TbRa2A with sera from *M. tuberculosis* patients and normal donors is shown in Table 4. The mean value for reactivity of TbRa2A with sera from *M. tuberculosis* patients was 0.444 with a standard deviation of 0.309. The mean for reactivity with sera from normal donors was 0.109 with a standard deviation of 0.029. Testing of 38 kD negative sera (Figure 7) also indicated that the TbRa2A antigen was capable of detecting sera in this category.

TABLE 4  
REACTIVITY OF TBRA2A WITH SERA FROM *M. TUBERCULOSIS* PATIENTS AND FROM  
NORMAL DONORS

Serum ID	Status	OD 450
Tb85	TB	0.680
Tb86	TB	0.450
Tb87	TB	0.263

Tb93	TB	0.232
Tb94	TB	0.333
Tb95	TB	0.435
Tb96	TB	0.284
Tb97	TB	0.320
Tb99	TB	0.328
Tb100	TB	0.817
Tb101	TB	0.607
Tb102	TB	0.191
Tb103	TB	0.228
Tb107	TB	0.324
Tb109	TB	1.572
Tb112	TB	0.338
DL4-0176	Normal	0.036
AT4-0043	Normal	0.126
AT4-0044	Normal	0.130
AT4-0052	Normal	0.135
AT4-0053	Normal	0.133
AT4-0062	Normal	0.128
AT4-0070	Normal	0.088
AT4-0091	Normal	0.108
AT4-0100	Normal	0.106
AT4-0105	Normal	0.108
AT4-0109	Normal	0.105

The reactivity of the recombinant antigen (g) (SEQ ID NO: 60) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. Figure 8 shows the results of the titration of antigen (g) with four *M. tuberculosis* positive sera that were all reactive with the 38 kD antigen and with four donor sera. All four positive sera were reactive with antigen (g).

The reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with sera from *M. tuberculosis* patients, PPD positive donors and normal donors was determined by indirect ELISA as described above. The results are shown in Figure 9. TbH-29 detected 30 out of 60 *M. tuberculosis* sera, 2 out of 8 PPD positive sera and 2 out of 27 normal sera.

Figure 9: Reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with sera from *M. tuberculosis* patients. The mean

OD 450 was demonstrated to be higher with sera from *M. tuberculosis* patients than from normal donors, with the mean OD 450 being significantly higher in the indirect ELISA than in the direct ELISA. Figure 11 is a titration curve for the reactivity of recombinant TbH-33 with sera from *M. tuberculosis* patients and from normal donors showing an increase in OD 450 with increasing concentration of antigen.

The reactivity of the recombinant antigens RDIF6, RDIF8 and RDIF10 (SEQ ID NOS: 184-187, respectively) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. RDIF6 detected 6 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; RDIF8 detected 14 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; and RDIF10 detected 4 out of 27 *M. tuberculosis* sera and 1 out of 15 normal sera. In addition, RDIF10 was found to detect 0 out of 5 sera from PPD-positive donors.

The antigens MO-1, MO-2, MO-4, MO-28 and MO-29 described above in Example 5, were expressed in *E. coli* and purified using a hexahistidine tag. The reactivity of these antigens with both *M. tuberculosis* positive and negative sera was examined by ELISA as described above. Titration curves showing the reactivity of MO-1, MO-2, MO-4, MO-28 and MO-29 at different solid phase coat levels when tested against four *M. tuberculosis* positive sera and four *M. tuberculosis* negative sera are shown in Figs. 12A-E, respectively. Three of the clones, MO-1, MO-2 and MO-29 were further tested on panels of HIV positive/tuberculosis (HIV TB) positive and extrapulmonary sera. MO-1 detected 5/20 extrapulmonary and 2/38 HIV TB sera. On the same sera groups, MO-2 detected 2/20 and 10/38, and MO-29 detected 2/20 and 8/38 sera. In combination these three clones would have detected 4/20 extrapulmonary sera and 16/38 HIV TB sera. In addition, MO-1 detected 6/17 sera that had previously been shown only to react with *M. tuberculosis* lysate and not with either 38 kD or with other antigens of the subject invention.

EXAMPLE 10PREPARATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was  
5 prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified  
by PCR in order to facilitate their fusion and the subsequent expression of the fusion  
protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform  
PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 141 and 142), PDM-57 and  
10 PDM-58 (SEQ ID NO: 143 and 144), and PDM-69 and PDM-60 (SEQ ID NO: 145-  
146), respectively. In each case, the DNA amplification was performed using 10 µl  
10X Pfu buffer, 2 µl 10 mM dNTPs, 2 µl each of the PCR primers at 10 µM  
concentration, 81.5 µl water, 1.5 µl Pfu DNA polymerase (Stratagene, La Jolla, CA)  
and 1 µl DNA at either 70 ng/µl (for TbRa3) or 50 ng/µl (for 38 kD and Tb38-1). For  
15 TbRa3, denaturation at 94°C was performed for 2 min, followed by 40 cycles of 96°C  
for 15 sec and 72°C for 1 min, and lastly by 72°C for 4 min. For 38 kD, denaturation at  
96°C was performed for 2 min, followed by 40 cycles of 96°C for 30 sec, 68°C for 15  
sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at 94°C  
for 2 min was followed by 10 cycles of 96°C for 15 sec, 68°C for 15 sec and 72°C for  
20 1.5 min, 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by  
72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned  
directly into pT7<sup>+</sup>L2IL-1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment  
was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and  
25 then digested with EcoRI for direct cloning into the pT7<sup>+</sup>L2Ra3-1 vector which was  
digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and  
EcoRI and directly subcloned into pT7<sup>+</sup>L2Ra3 38kD-17 digested with the same  
enzymes. The whole fusion was then transferred to pET28b (Novagen, Madison, WI) by

The expression construct was transformed to BLR pLys S *E. coli* (Novagen, Madison, WI) and grown overnight in LB broth with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD<sub>560</sub> of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 µg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26,000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0), 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen, Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialyzed against 10 mM Tris (8.0).

The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbRa3-38 kD-Tb38-1) are provided in SEQ ID NO: 147 and 148, respectively.

A fusion protein containing the two antigens TbH-9 and Tb38-1 (hereinafter referred to as TbH9-Tb38-1) without a hinge sequence, was prepared using a similar procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 151.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-59 (SEQ ID NO: 145) and PDM-62 (SEQ ID NO: 200) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a DraI site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3 38kD Tb38-1A fusion was then transferred to pET28b using NdeI and EcoRI sites.

Denaturation at 94 °C was performed for 2 min, followed by 10 cycles of 96 °C for 15 sec, 68 °C for 15 sec and 72 °C for 1.5 min; 30 cycles of 96 °C for 15 sec, 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with EcoRI and Eco72I and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-2) are provided in SEQ ID NO: 203 and 204, respectively.

10 A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and TbH4 was prepared as follows.

Genomic *M. tuberculosis* DNA was used to PCR full-length TbH4 (FL TbH4) with the primers PDM-157 and PDM-160 (SEQ ID NO: 343 and 344, respectively) and 2 µl DNA at 100 ng/µl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 61 °C for 20 sec and 72 °C for 5 min; and finally by annealing at 72 °C for 10 min. The FL TbH4 PCR fragment was digested with EcoRI and Sca I (New England Biolabs.) and cloned directly into the pET28Ra3/38kD/38-1A construct described above which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-6) are provided in SEQ ID NO: 345 and 346, respectively.

A fusion protein containing the antigen 38kD and DPEP separated by a linker was prepared as follows.

25 38 kD DNA was used to perform PCR using the primers PDM-176 and PDM-175 (SEQ ID NO: 347 and 348, respectively), and 1 µl PET28Ra3/38kD/38-1Ra2A-12 DNA at 110 ng/µl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 72 °C for 15 sec and 72 °C for 5 min.

then ramping down to 25 °C slowly at 0.1 °C/sec. DPEP DNA was used to perform PCR as described above. The 38 kD fragment was digested with Eco RI (New England Biolabs) and cloned into a modified pT7ΔL2 vector which was cut with Eco 72 I (Promega) and Eco RI. The modified pT7ΔL2 construct was designed to have a 5 MGHHHHHH amino acid coding region in frame just 5' of the Eco 72 I site. The construct was digested with Kpn 2I (Gibco, BRL) and Pst I (New England Biolabs) and the annealed sets of phosphorylated primers (PDM-171, PDM-172 and PDM-173, PDM-174) were cloned in. The DPEP PCR fragment was digested with Eco RI and Eco 72 I and cloned into this second construct which was digested with Eco 47 III (New England Biolabs) and Eco RI. Ligations were done with a ligation kit from Panvera (Madison, WI). The resulting construct was digested with NdeI (New England Biolabs) and Eco RI, and transferred to a modified pET28 vector. The fusion construct was confirmed to be correct by DNA sequencing.

Recombinant protein was prepared essentially as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-8) are provided in SEQ ID NO: 349 and 350, respectively.

#### EXAMPLE 11

##### USE OF M. TUBERCULOSIS FUSION PROTEINS FOR SERODIAGNOSIS OF TUBERCULOSIS

The effectiveness of the fusion protein TbRa3-38 and Tb38-1, prepared as described above, in the serodiagnosis of tuberculosis infection was examined by ELISA.

The ELISA protocol was as described above in Example 6, with the fusion protein being coated at 200 ng/well. A panel of sera was chosen from a group of tuberculosis patients previously shown, either by ELISA or by culture, to be

all three epitopes functioned with the fusion protein. As shown in Table 5, all four sera that reacted with TbRa3 only were detectable with the fusion protein. Three sera that reacted only with Tb38-1 were also detectable, as were two sera that reacted with 38 kD alone. The remaining 15 sera were all positive with the fusion protein based on a cut-off in the assay of mean negatives +3 standard deviations. This data demonstrates the functional activity of all three epitopes in the fusion protein.

TABLE 5  
REACTIVITY OF TRI-PEPTIDE FUSION PROTEIN WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Serum ID	Status	ELISA and/or Western Blot Reactivity with Individual proteins			Fusion Recombinant OD 450	Fusion Recombinant Status
		38kd	Tb38-1	TbRa3		
01B93I-40	TB	-	-	-	0.413	-
01B93I-41	TB	-	-	-	0.392	-
01B93I-29	TB	-	-	-	2.217	-
01B93I-109	TB	-	-	-	0.522	-
01B93I-132	TB	-	-	-	0.937	-
5004	TB	-	-	-	1.098	-
15004	TB	-	-	-	2.077	-
39004	TB	-	-	-	1.675	-
68004	TB	-	-	-	2.388	-
99004	TB	-	-	-	0.607	-
107004	TB	-	-	-	0.667	-
92004	TB	-	-	-	1.070	-
97004	TB	-	-	-	1.152	-
118004	TB	-	-	-	2.694	-
173004	TB	-	-	-	3.258	-
175004	TB	-	-	-	2.514	-
274004	TB	-	-	-	-	-

308004	TB	-	-	-	3.338	+
314004	TB	-	+	-	1.362	+
317004	TB	-	-	-	0.763	+
312004	TB	-	-	+	1.079	+
D176	PPD	-	-	-	0.145	-
D162	PPD	-	-	-	0.073	-
D161	PPD	-	-	-	0.097	-
D27	PPD	-	-	-	0.082	-
A6-124	NORMAL	-	-	-	0.053	-
A6-125	NORMAL	-	-	-	0.087	-
A6-126	NORMAL	-	-	-	0.346	-
A6-127	NORMAL	-	-	-	0.064	-
A6-128	NORMAL	-	-	-	0.034	-
A6-129	NORMAL	-	-	-	0.037	-
A6-130	NORMAL	-	-	-	0.057	-
A6-131	NORMAL	-	-	-	0.054	-
A6-132	NORMAL	-	-	-	0.022	-
A6-133	NORMAL	-	-	-	0.147	-
A6-134	NORMAL	-	-	-	0.101	-
A6-135	NORMAL	-	-	-	0.066	-
A6-136	NORMAL	-	-	-	0.054	-
A6-137	NORMAL	-	-	-	0.065	-
A6-138	NORMAL	-	-	-	0.041	-
A6-139	NORMAL	-	-	-	0.103	-
A6-140	NORMAL	-	-	-	0.212	-
A6-141	NORMAL	-	-	-	0.156	-
A6-142	NORMAL	-	-	-	0.051	-

The reactivity of the fusion protein TbF-2 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 6) demonstrate that all four antigens function independently in the ELISA assay.

TABLE 6  
REACTIVITY OF TbF-2 FUSION PROTEIN WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status	ELISA Reactivity			
						38 kD	TbRa3	Tb38-1	DPEP
B931-40	TB	0.57	-	0.321	+	-	-	-	-
B931-41	TB	0.601	+	0.396	-	+	+	-	-
B931-109	TB	0.494	-	0.404	+	+	-	-	-
B931-132	TB	1.502	-	1.292	-	-	-	-	-
5004	TB	1.806	-	1.666	-	-	-	-	-
15004	TB	2.862	-	2.468	-	-	-	-	-
39004	TB	2.443	-	1.722	-	-	-	-	-
68004	TB	2.871	-	2.575	-	-	-	-	-
99004	TB	0.691	-	0.971	-	-	-	-	-
107004	TB	0.875	-	0.732	-	-	-	-	-
92004	TB	1.632	-	1.394	-	-	-	-	-
97004	TB	1.491	-	1.979	+	-	-	-	-
118004	TB	3.182	-	3.045	+	-	-	-	-
173004	TB	3.644	-	3.578	-	-	-	-	-
175004	TB	3.332	-	2.916	-	-	-	-	-
274004	TB	3.696	-	3.716	-	-	-	-	-
276004	TB	3.243	-	2.56	-	-	-	-	-
282004	TB	1.249	-	1.234	-	-	-	-	-
289004	TB	1.373	-	1.17	-	-	-	-	-
308004	TB	3.708	-	3.355	-	-	-	-	-
314004	TB	1.663	-	1.290	-	-	-	-	-
317004	TB	1.163	-	0.92	-	-	-	-	-
312004	TB	1.709	-	1.453	-	-	-	-	-
380004	TB	0.228	-	0.461	-	-	-	-	-
451004	TB	0.18	-	0.2	-	-	-	-	-
475004	TB	0.188	-	0.469	-	-	-	-	-
410004	TB	0.384	-	2.392	-	-	-	-	-
411004	TB	0.306	-	0.874	-	-	-	-	-
421004	TB	0.337	-	1.456	-	-	-	-	-
428004	TB	0.047	-	0.396	-	-	-	-	-
A6-8	Normal	0.094	-	0.093	-	-	-	-	-
A6-88	Normal	0.214	-	0.19	-	-	-	-	-
A6-89	Normal	0.248	-	0.125	-	-	-	-	-
A6-90	Normal	0.179	-	0.206	-	-	-	-	-
A6-91	Normal	0.138	-	0.151	-	-	-	-	-
A6-92	Normal	0.064	-	0.097	-	-	-	-	-
A6-93	Normal	0.072	-	1.098	-	-	-	-	-
A6-94	Normal	0.072	-	0.064	-	-	-	-	-
A6-95	Normal	0.125	-	0.159	-	-	-	-	-
A6-96	Normal	0.072	-	0.064	-	-	-	-	-

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the  
5 construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the  
10 invention.

## CLAIMS

We claim:

1. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or in modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);
- (b) Ala-Val-Glu-Ser-Gly-Met-Ile-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 17);
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123); and
- (j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly (SEQ ID NO: 131)

wherein Xaa may be any amino acid

2. A polypeptide comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124) and
- (b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any amino acid.

3. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an antigenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-175, 196, 235, 237-242, 248-251, 290-293, 304, 311, 313-315, 317, 319, 323, 324, 328, 330, 332, 334 and 336, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-175, 196, 235, 237-242, 248-251, 290-293, 304, 311, 313-315, 317, 319, 323, 324, 328, 330, 332, 334 and 336, or a complement thereof under moderately stringent conditions.

6. A recombinant expression vector comprising a DNA molecule according to claim 5.
7. A host cell transformed with an expression vector according to claim 6.
8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.
9. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:
  - (a) contacting a biological sample with one or more polypeptides according to any of claims 1-4; and
  - (b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.
10. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:
  - (a) contacting a biological sample with a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and
  - (b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.
11. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:
  - (a) contacting a biological sample with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NO: 129 and 130;

sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

12. The method of any one of claims 9-11 wherein step (a) additionally comprises contacting the biological sample with a 38 kD *M. tuberculosis* antigen and step (b) additionally comprises detecting in the sample the presence of antibodies that bind to the 38 kD *M. tuberculosis* antigen.

13. The method of any one of claims 9-11 wherein the polypeptide(s) are bound to a solid support.

14. The method of claim 13 wherein the solid support comprises nitrocellulose, latex or a plastic material.

15. The method of any one of claims 9-11 wherein the biological sample is selected from the group consisting of whole blood, serum, plasma, saliva, cerebrospinal fluid and urine.

16. The method of claim 15 wherein the biological sample is whole blood or serum.

17. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a

(b) detecting in the sample a DNA sequence that amplifies in the presence of the oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

18. The method of claim 17, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA molecule according to claim 5.

19. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337; and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

20. The method of claim 19, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

21. The method of claims 17 or 19 wherein the biological sample selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

22. A method for detecting *M. tuberculosis* infection in a biological

- (a) contacting the sample with one or more oligonucleotide probes specific for a DNA molecule according to claim 5; and
- (b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

23. The method of claim 22 wherein the probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

24. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

- (a) contacting the sample with one or more oligonucleotide probes specific for a DNA sequence selected from the group consisting of SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337; and
- (b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

25. The method of claim 24 wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

26. The method of claims 22 or 24 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

27. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide according to any one of claims 1-4; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

28. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO 129 and 130; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

29. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

30. The method of any one of claims 27-29 wherein the binding agent is a monoclonal antibody.

31. The method of any one of claims 27-29 wherein the binding agent is a polyclonal antibody.
32. A diagnostic kit comprising:
- (a) one or more polypeptides according to any of claims 1-4; and
  - (b) a detection reagent.
33. A diagnostic kit comprising:
- (a) one or more polypeptides having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and
  - (b) a detection reagent.
34. A diagnostic kit comprising:
- (a) one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337, and
  - (b) a detection reagent.
35. The kit of any one of claims 32-34 wherein the polypeptides are immobilized on a solid support.
36. The kit of claim 35 wherein the solid support comprises nitrocellulose, latex, or a plastic material.

38. The kit of claim 37 wherein the binding agent is selected from the group consisting of anti-immunoglobulins, Protein G, Protein A and lectins.
39. The kit of claim 37 wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin, dye particles and colloidal particles.
40. A diagnostic kit comprising at least two oligonucleotide primers, at least one of the oligonucleotide primers being specific for a DNA molecule according to claim 5.
41. A diagnostic kit according to claim 40, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA molecule according to claim 5.
42. A diagnostic kit comprising a at least two oligonucleotide primers, at least one of the primers being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.
43. A diagnostic kit according to claim 42, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

45. A kit according to claim 44, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

46. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

47. A kit according to claim 46, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195, 198, 210-220, 232, 234, 256-271, 287, 288, 298-303, 305-310, 312, 316, 318, 320-322, 325-327, 329, 331, 333, 335 and 337.

48. A monoclonal antibody that binds to a polypeptide according to any of claims 1-4.

49. A polyclonal antibody that binds to a polypeptide according to any of claims 1-4.

50. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.

51. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6 (SEQ ID NO 99).

53. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 38 kD (SEQ ID NO: 150).

54. A diagnostic kit comprising:

- (a) one or more fusion proteins according to any one of claims 50-53; and
- (b) a detection reagent.

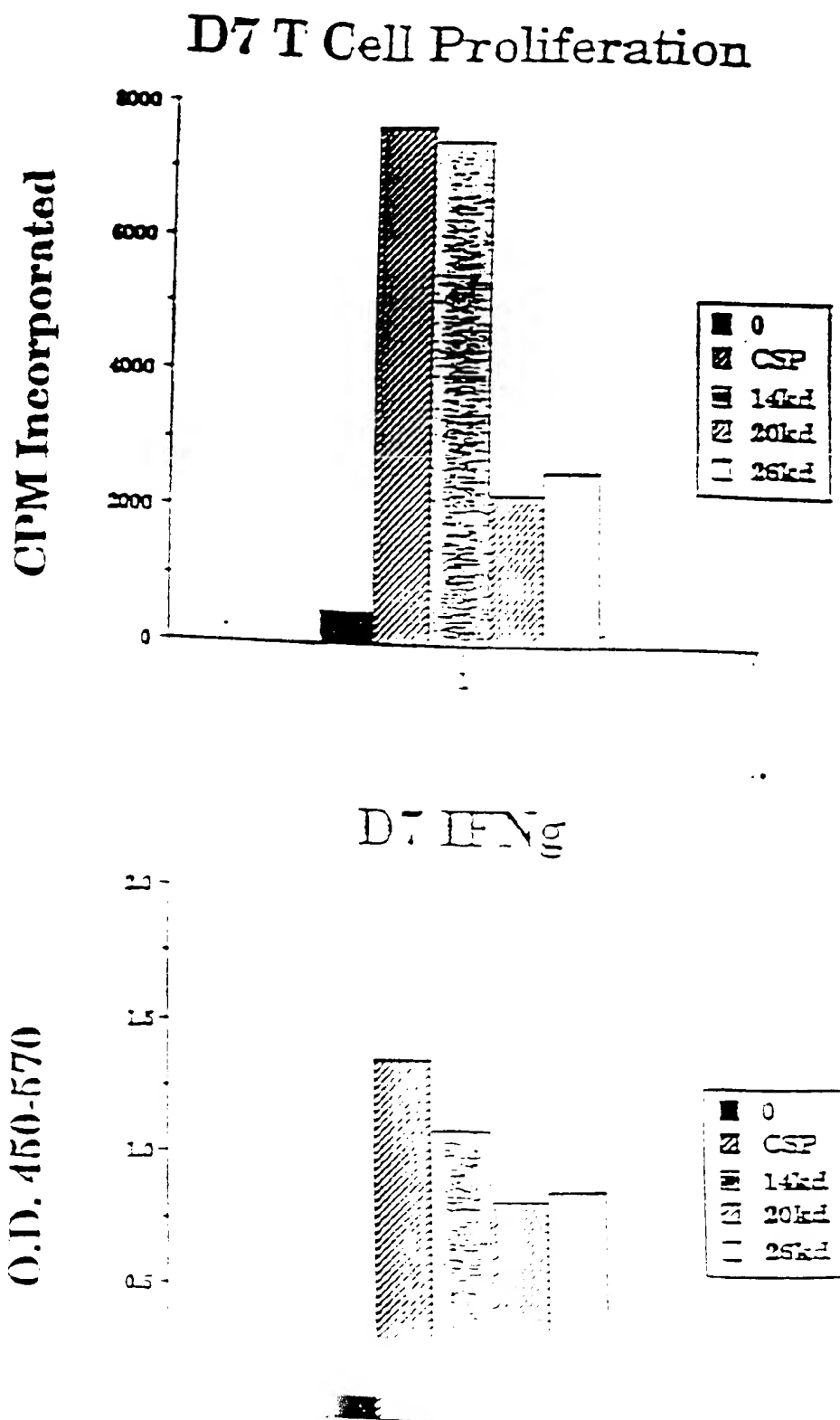
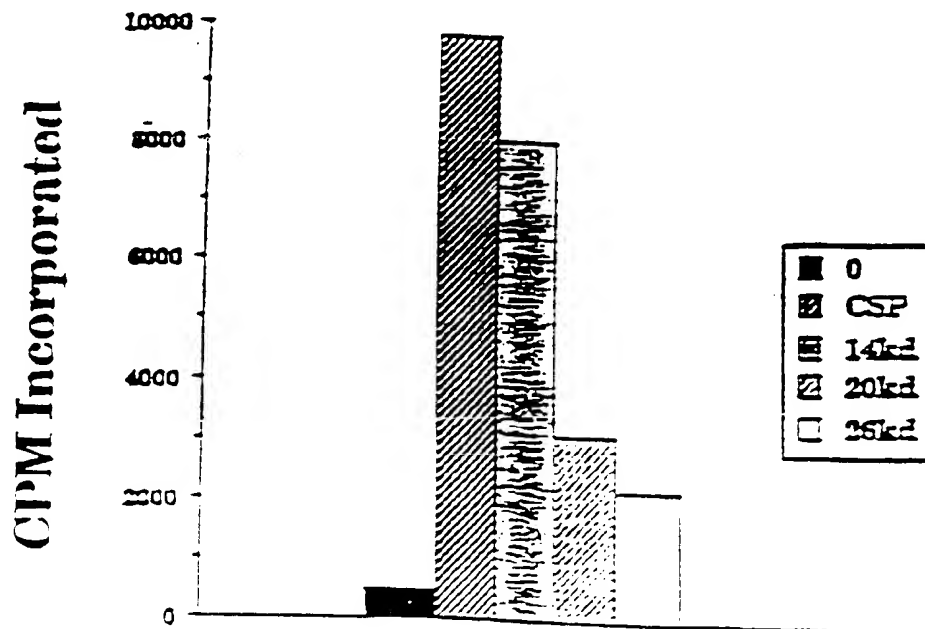
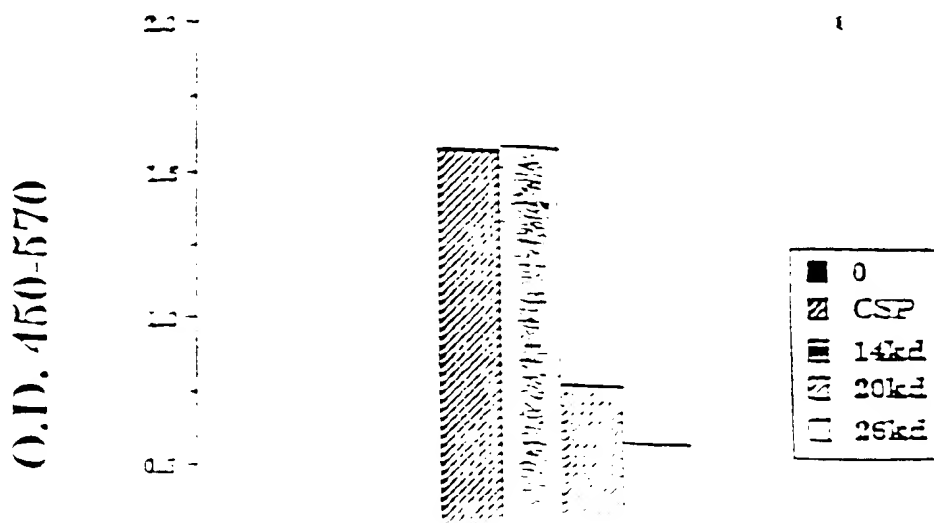


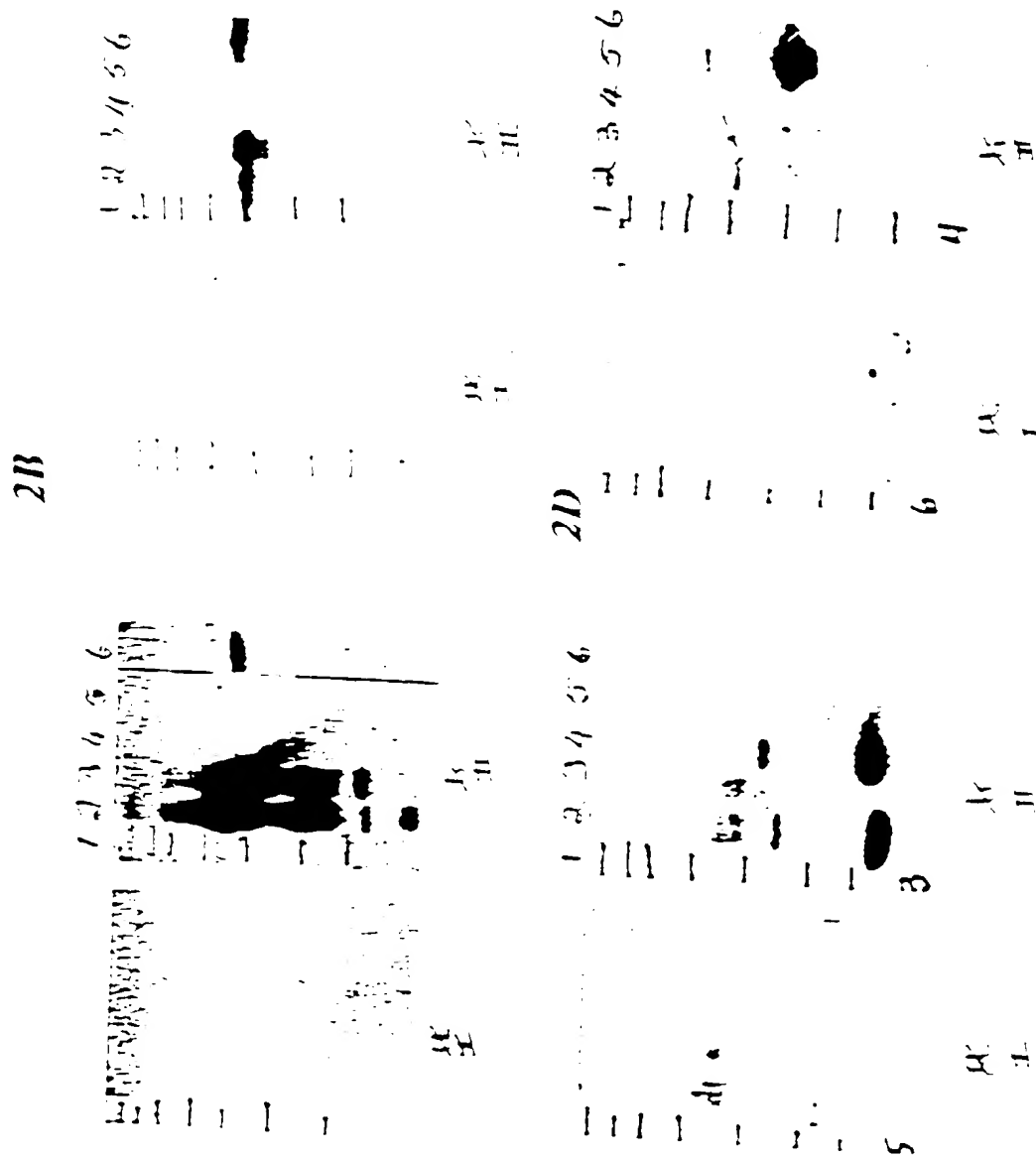
FIG. 1A

## D160 T Cell Proliferation



## D160 IFN<sub>γ</sub>





FIGS. 2A-D

# T cell clone 131TbH9 responds poorly to CSP

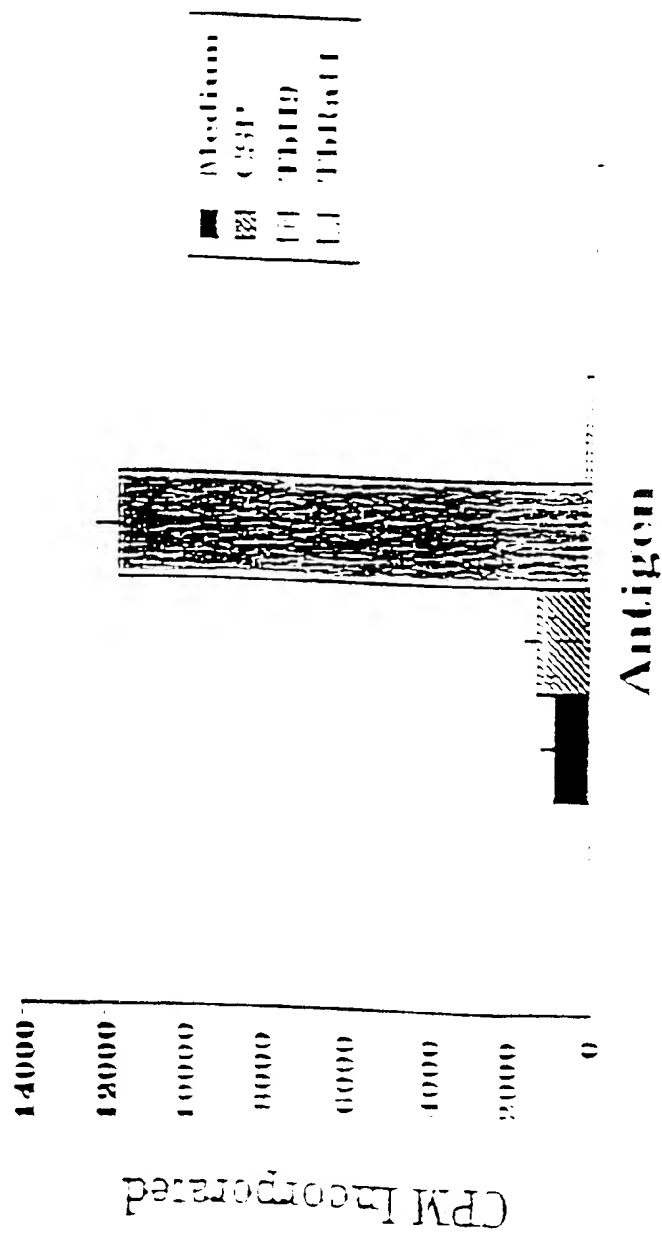


FIG. 3A

# T Cell Clone PPD 800-10 IFN $\gamma$ Production

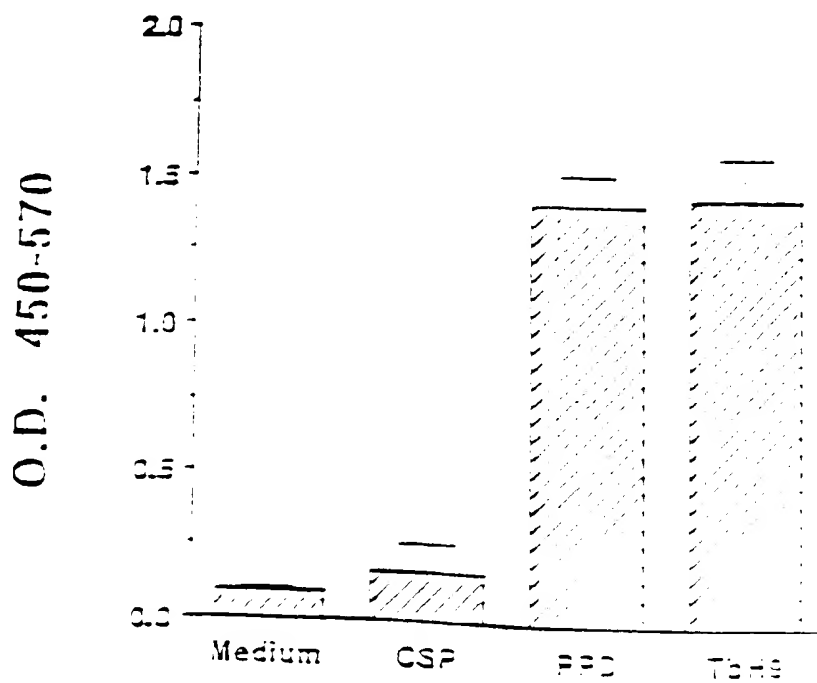


FIG. 3B

Anti-Genotoxin Antibody Titration

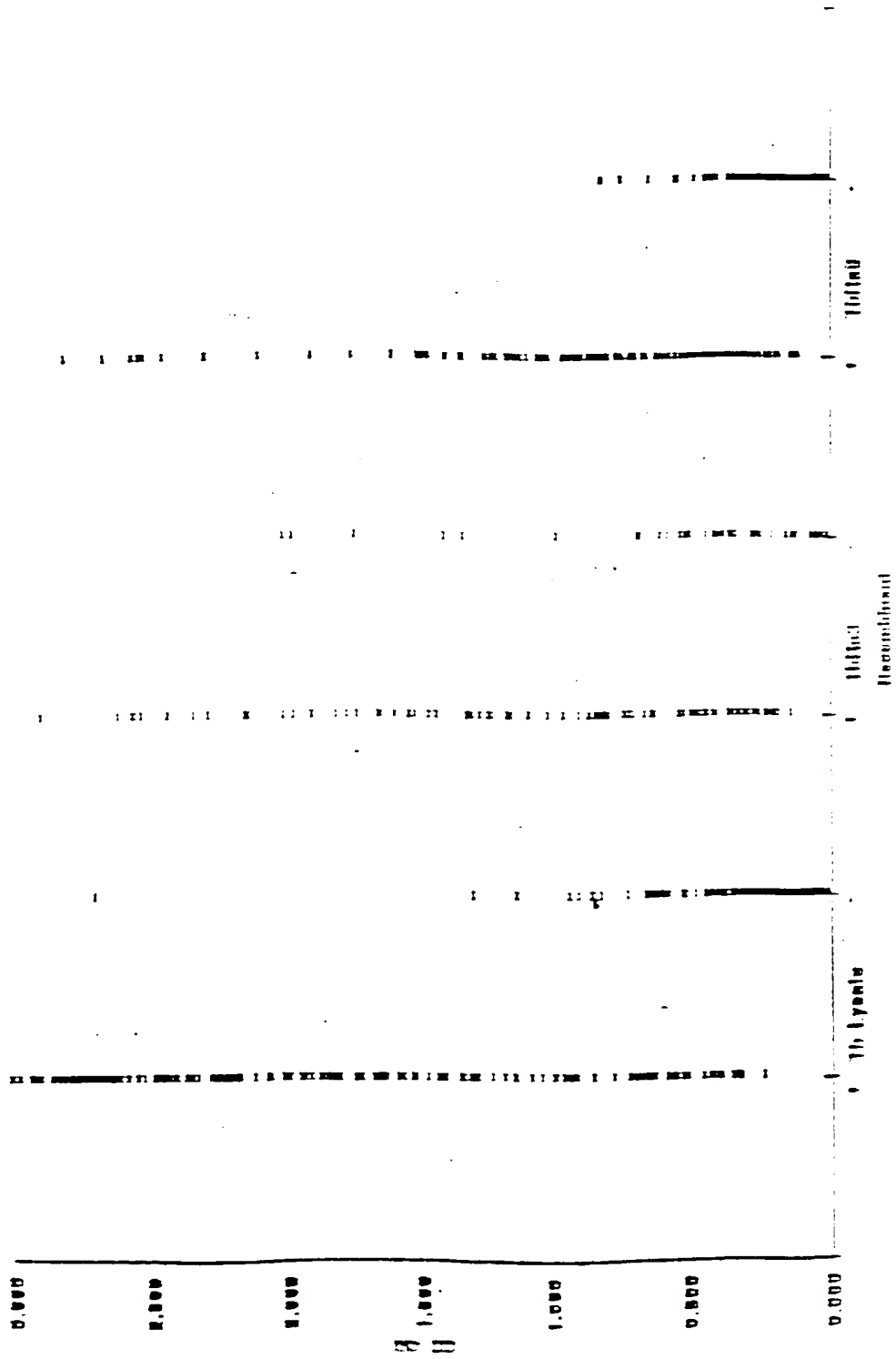


FIG. 3

Relative Sensitivity of Fluorescence and Peptides obtained using Human T0 Bone marrow procedure

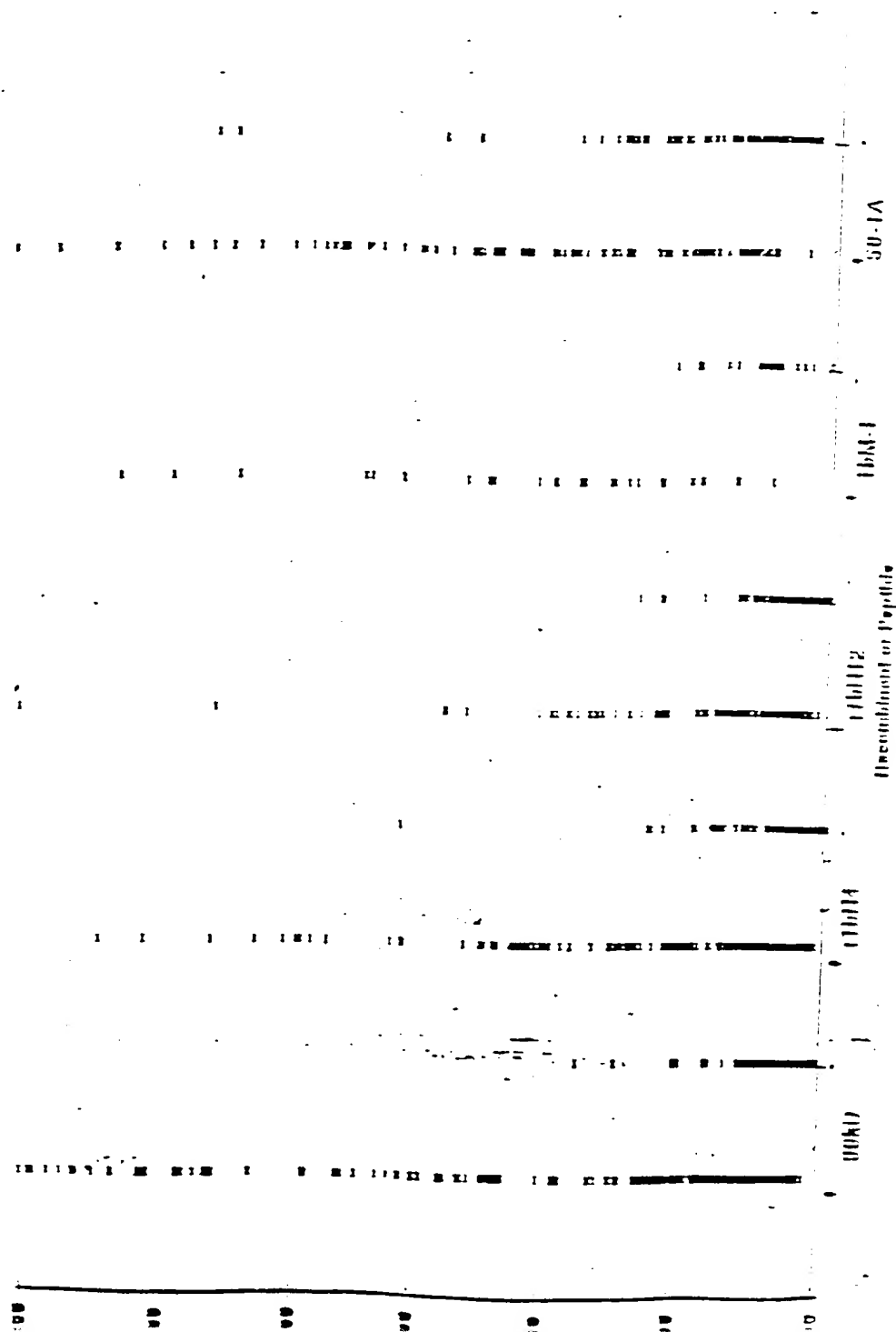


FIG. 6

Reactivity of Recombinant 30kD and 100kD antigens with sera from *M. tuberculosis* patients, PPD positive and normal donors

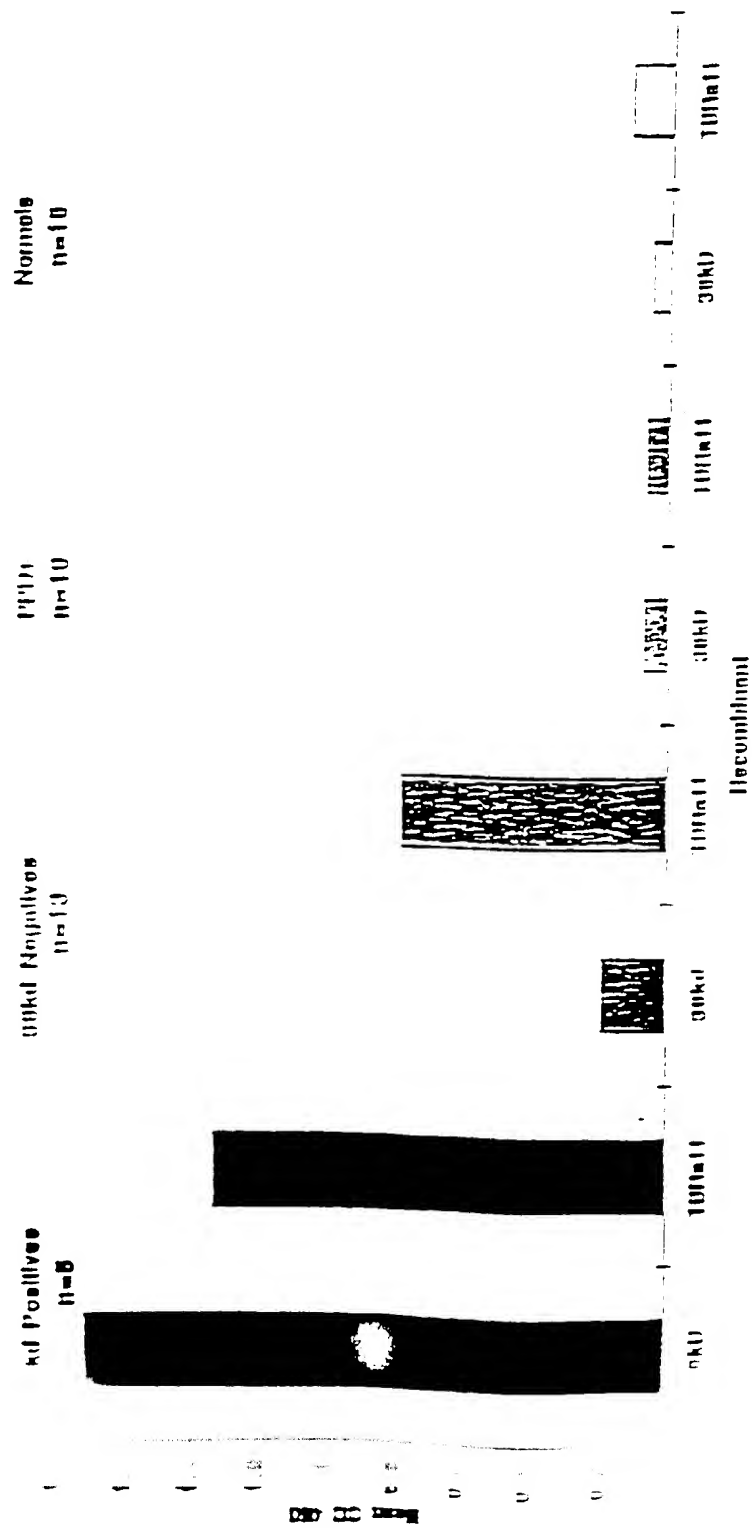
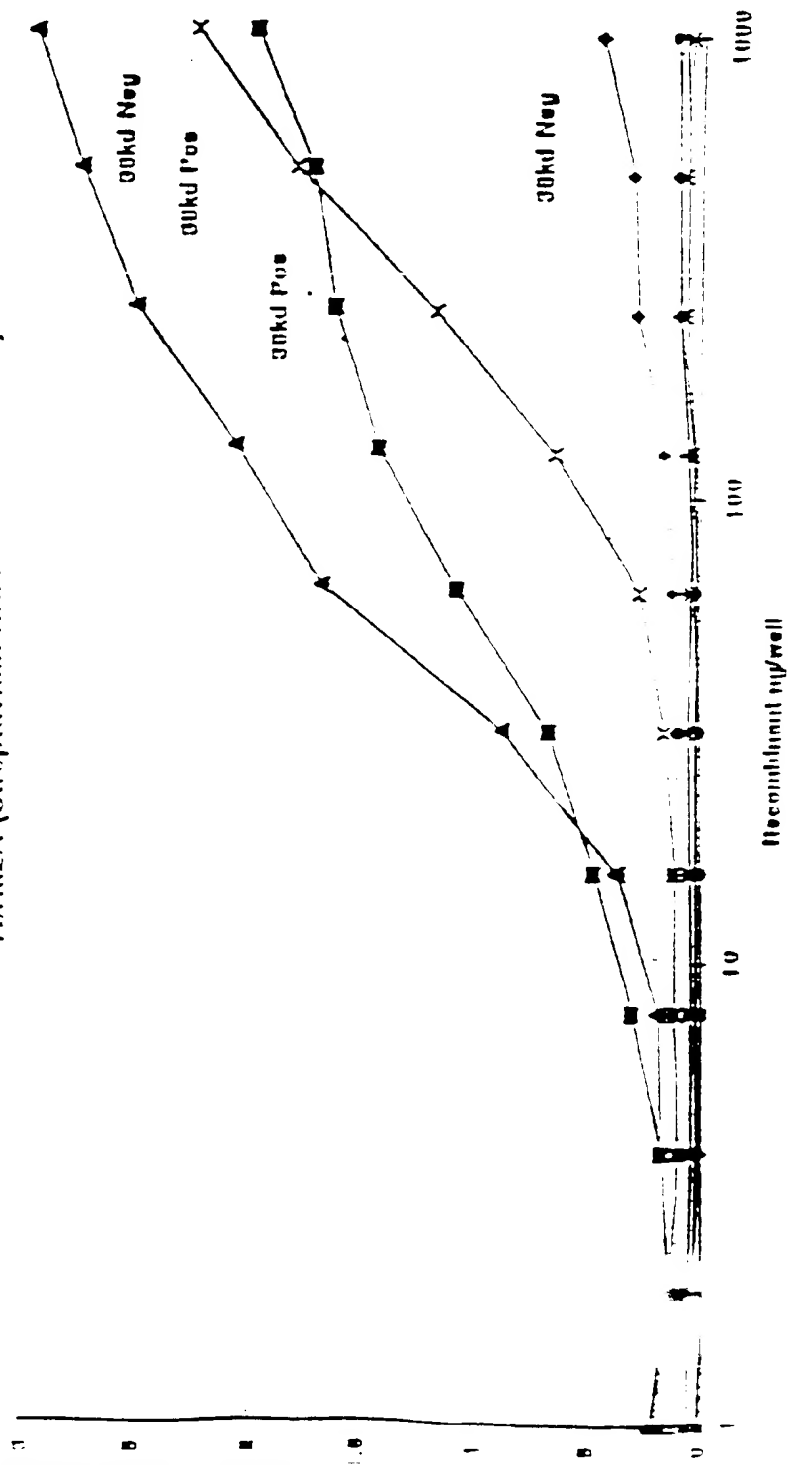


FIG. 7

TbHte2A (Streptococcus HRP/Protein A Blot)



◆	10004
■	67004
▲	60004
×	30004
◆	A114
●	A110
+	A140
△	A141

Activity of DPEP recombinant with 10 positive and negative sera

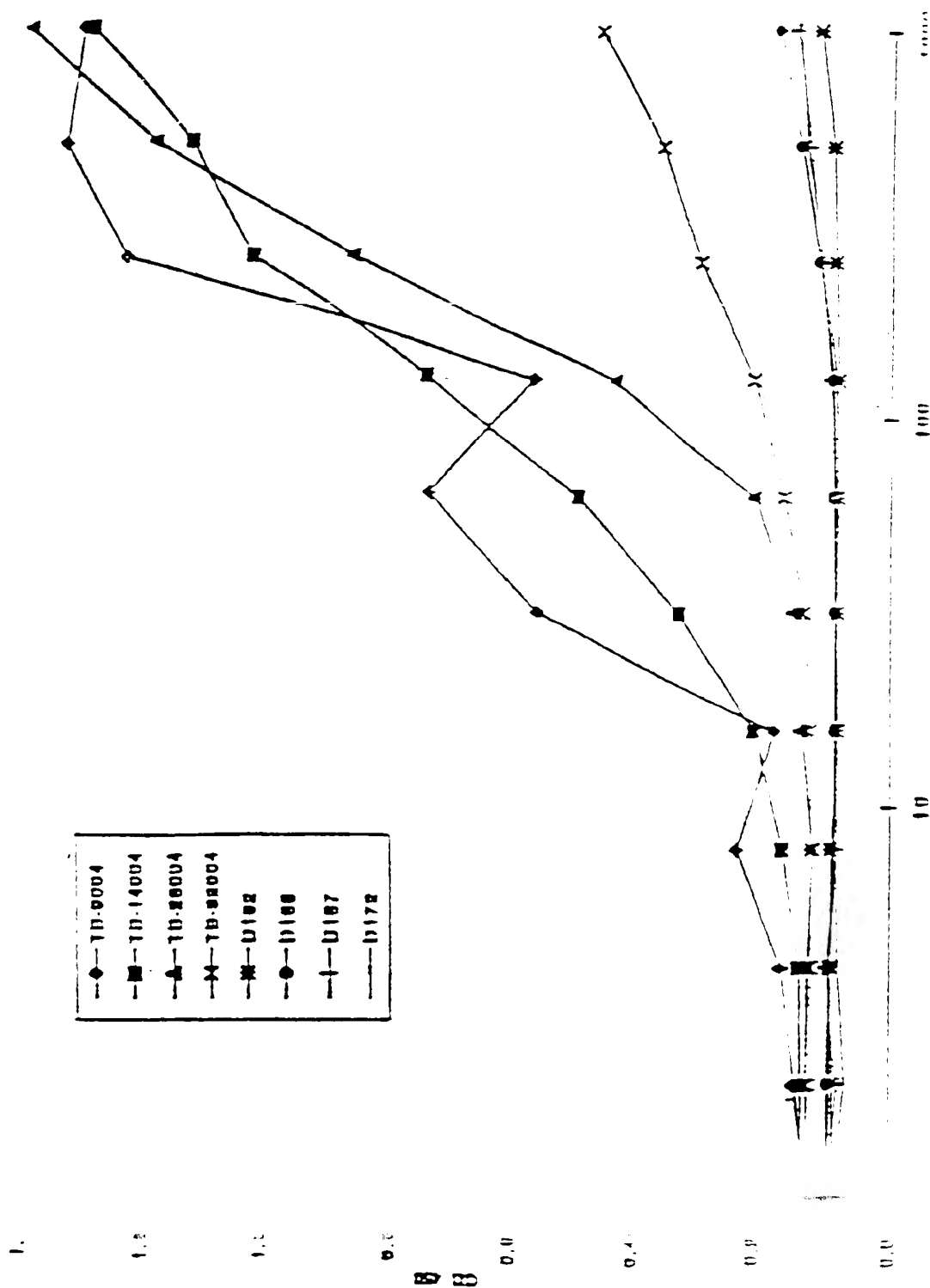
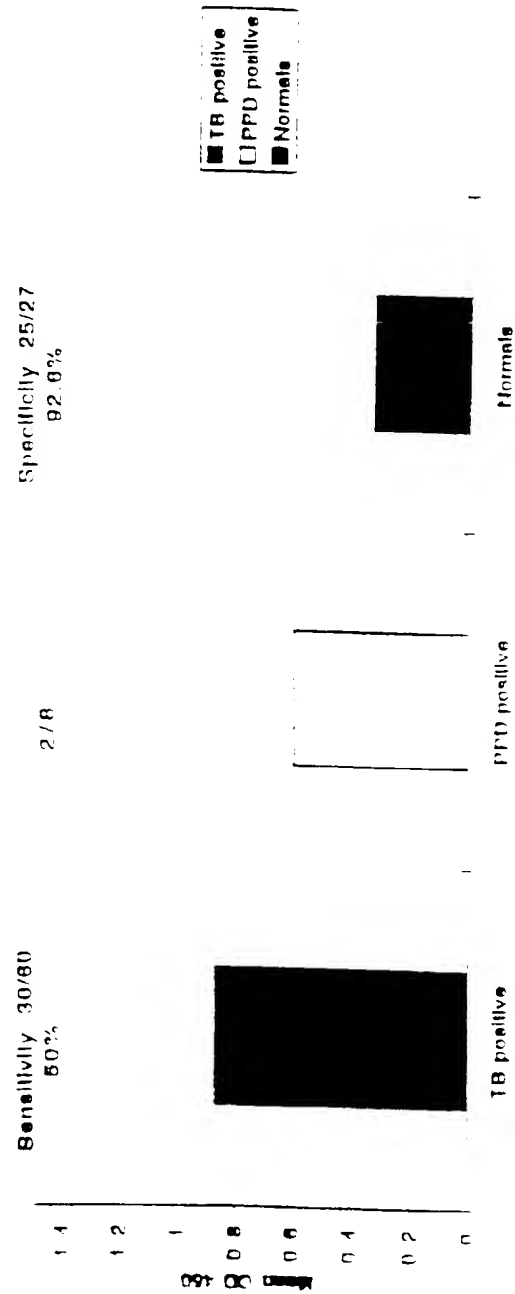
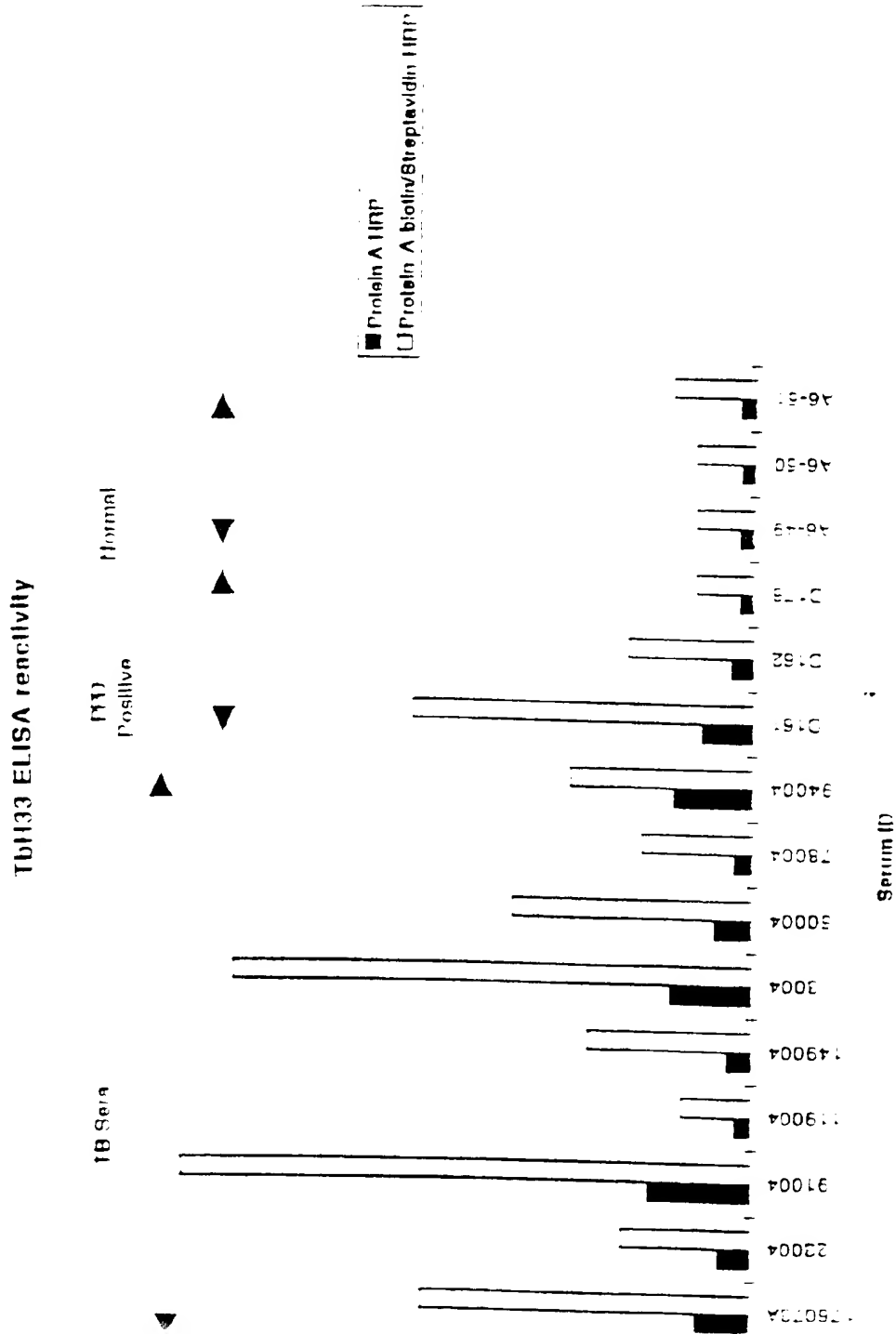


FIG. 9

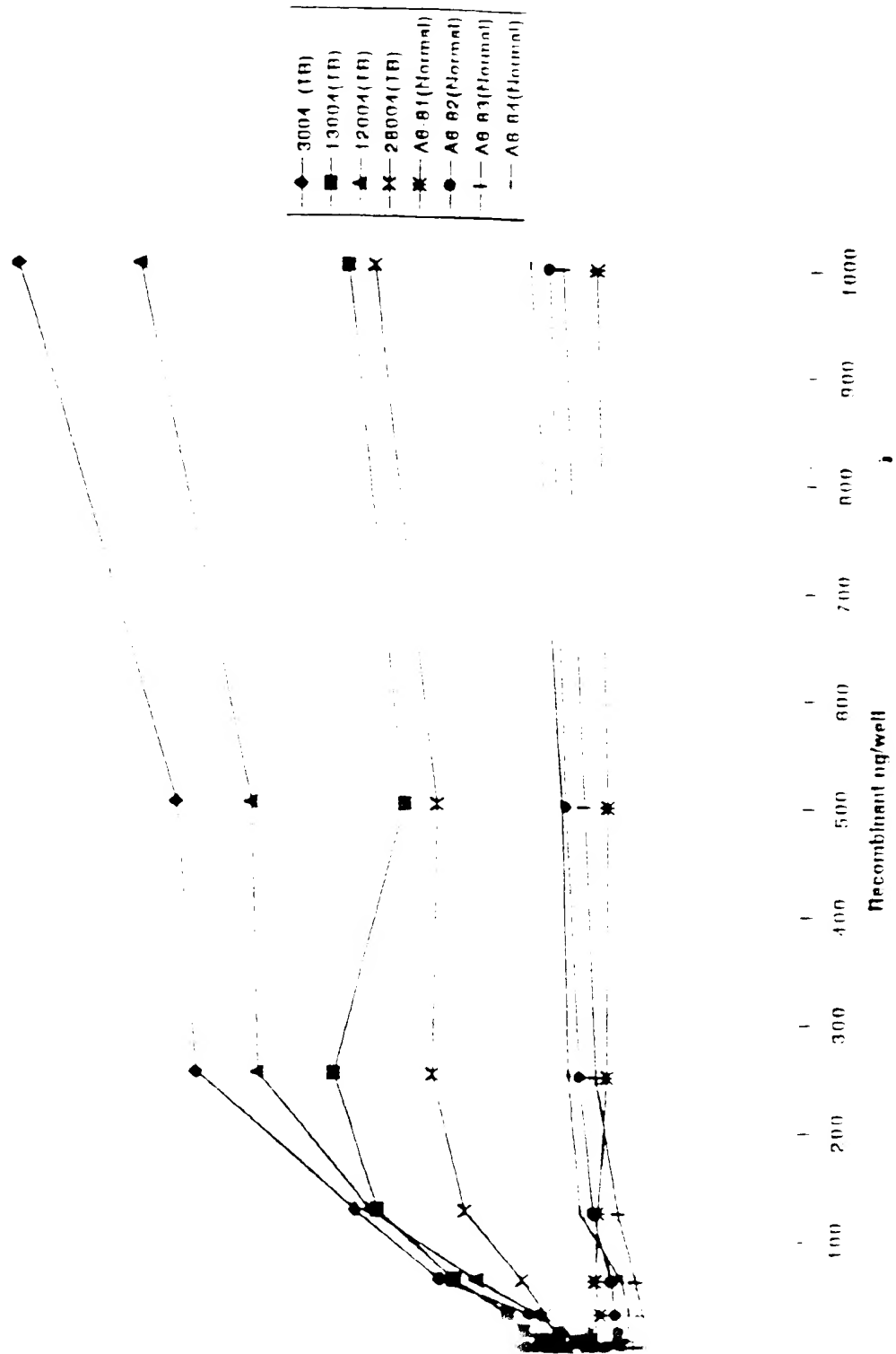
TbH29 ELISA reactivity in Streptavidin-HRP/Protein A-biotin system



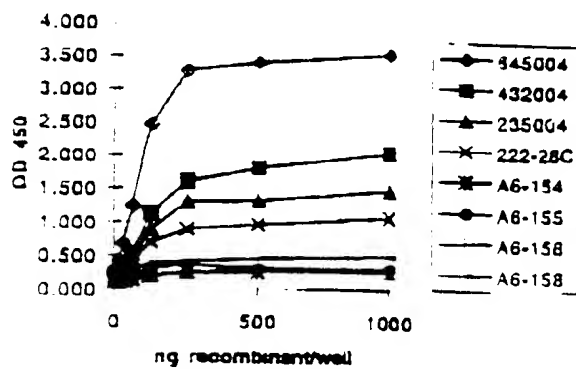


**FIG. 10**

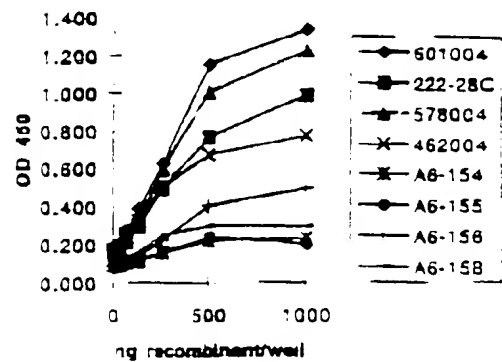
ELISA reactivity of Tblt33



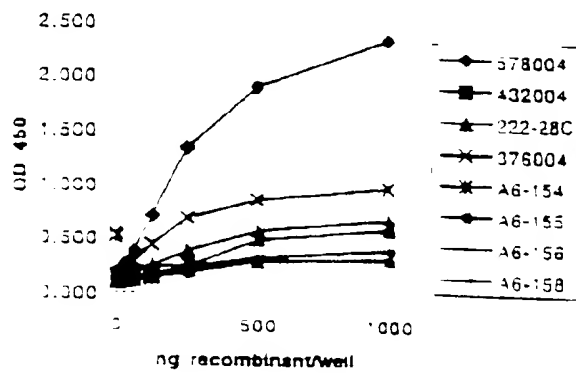
Titration of Mo-1 antigen with TB  
positive and negative sera



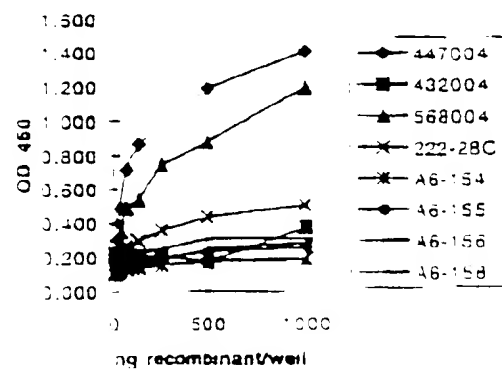
Titration of Mo-2 with TB  
positive and negative sera



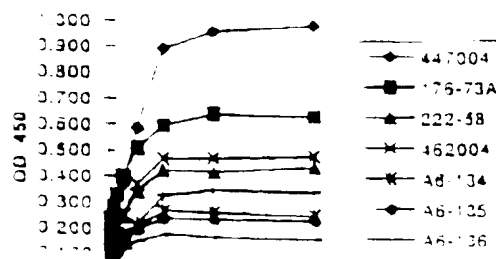
Titration of Mo-4 with TB positive  
and negative sera



Titration of Mo-28 with TB  
positive and negative sera



Titration of Mo-29 with TB  
positive and negative sera



## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(1) APPLICANTS: Reed, Steven G.  
Skeiky, Yasir A.W.  
Dillon, Davin C.  
Campos-Neto, Antonia  
Houghton, Raymond  
Vedvick, Thomas S.  
Twardzik, Daniel R.  
Lodes, Michael J.  
Hendrickson, Ronald

(2) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
TUBERCULOSIS

(3) NUMBER OF SEQUENCES: 300

## (4) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SEED and BERRY LLP  
(B) STREET: 6300 Columbia Center, 701 Fifth Avenue  
(C) CITY: Seattle  
(D) STATE: Washington  
(E) COUNTRY: USA  
(F) ZIP: 98104-7092

## (5) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.23, Version #1.10

## (6) IMPRINT APPLICATION DATA:

(A) APPLICATION NUMBER  
(B) FILING DATE: 15-MAY 1998  
(C) CLASSIFICATION:

## (7) ATTORNEY/AGENT INFORMATION:

(A) NAME: Marki, David  
(B) REGISTRATION NUMBER: 1000  
(C) REFERENCE NUMBER: 1000000000

## (8) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 206 422 4500  
(B) TELEFAX: 206 482 6031

INFORMATION FOR SEQ. ID NO. 1

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGAGGCACCG GTAGTTTGAA CCAACGCAC AATCGACGG CAAACGAAC GAAGAACACA	50
ACCATGAAGA TGGTGAAATC GATCGCCGCA GGTCTGACCG CCGCGGCTGC AATCGGCGCC	120
GCTGCGGCGG GTCTGACTTC GATCATGGCT GCGGCGCGG TCGTATACCA GATGCAGCCG	180
GTCTCTTTGG GCGCGGCACT GCGCTTGGAC CCGGCATCCG CCGCTGACCT CCGGACCGCC	240
GCGGAGTTGA CGAGCCTGCT CAACAGCCTG GCGGATCCCA ACCTGTCTTT TCGGAACAAG	300
GCGAGTCTGG TCGAGGCGCG CATCGGGGGC ACCGAGGCGC GCATCGCGCA CCACAAGCTG	360
AACAAGCGCG CCGAGCACCG GCATCTCGCG CTCTCTTTCA GCTGACGAA GATCCAGCCG	420
CGCGCGCGCG CTTCGCGGAC CCGCGAGCTT TCGCTCTCGG GTCGGAAGCT GTCTCGCGCG	480
GTGAGCGAGA ACCTCAGCTT GCTGAATCAA GCGCGCTGGA TGCTGTCACG CCGATCGCGC	540
ATGAGCTTGG TCGAGGCGCG AGGCAACTG ATTGGCGGGC CGGNTTCAGC CGCTCTTTCA	600
GCTACCGCGC CCGCCTGCTG ACCGCTCCAT GTCGAAGACT CCGCGGTGTA GCACGCTGCG	660
CTCTCGCGCG GCGCGCGCG ACCGCGCGCT GCAAGCGCTG CTGAGATAG CTGCTGCTG	720
GCGAGCGAG AGCAGCGCGG GCGCGCGCT TCTGCTGCTG GATGGA	780

(X2) INFORMATION FOR SEQ ID NO:1:

- (A) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 789 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X3) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGATGATG ATCAAGTCA ATATGAATG AGATAGAGA GAACTGCTT CTCTCTCTTA	120
CACTCTCTTA CTGAGCTTGA CTCTCTCTT CAAGCTCTTA CTGAGAGCT CTCTCTCTTA	240
CTGAGAGCTT TCGCGCGCGG CTCTCTCTTA CTCTCTCTTA AAGGAGGCGG CAAGCTCTTA	360
CTCTCTCTTA TACTGAGCA AGGCTCTCTT CTCTCTCTTA GCGATCGCGA CAGGAGCTTA	480
CTCTCTCTTA AGGCTCTCTT AGGCTCTCTT CTCTCTCTTA CTCTCTCTTA AAGGAGCTTA	600
CTCTCTCTTA CTCTCTCTTA CTCTCTCTTA CTCTCTCTTA CTCTCTCTTA CTCTCTCTTA	720

GCGCGCGCGCT GGCGCGGATG TCGATCGGGG CGGTCTCTCG ACCTGCTACG ACCGGATTTT 540  
 CGCTGATGTC CACCATCTCC AAGATTGAT TCTTGGGAGG CTTGAGGGTC NGGGTGACCC 600  
 CCGCGCGGGG CTCATTGCGG GGTNTCGGCN GGTTCACCC CTTACCNACT GCCNCCCGGN 660  
 TTGCNAATTC NTTCTTCTCT GCGCNAAAAG GGACCNNTAN CTTGCGCGCTN GAAANGGTNA 720  
 TCCNGGGGCCC NTCTTNGAAN CCGCNTCCCC CT 752

2) INFORMATION FOR SEQ ID NO:3:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 813 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

2) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CATATGCATC ACCATCACCA TCACATCTCT AACCGCCGAG CGCGTGGGGG GCGTCAGCCA 60  
 CCACCGGACA CCGGGCCCCG TCGATCTCTT AGCTTGAGTC TGGTCAGGCA TCGTCGTCAG 120  
 CACCGCGATG CCGTATGTTT GTCTGCAAT CAGATATCGC GGCAATCCAA TCTCCCGCCT 180  
 CCGCGCGGGG CTCTGCAAA CTACTCGGG AGGAATTTTC AGCTCCCAT CAAGATCTTC 240  
 ATCTGCTCA CCGCTCTCTT TTTGCTCTCT TCTTCGGGTC TGGCCACCGC CGGCGCCAG 300  
 AGCTATCTCC AGGATTTGAA AGGCACCAT ACCCGCCAGG CTGCGCAGAT TCAATCTTC 360  
 CACCTCTCTT ACAACATCAA CATCAGCTCT CCAATTTACT ACCCGACCA TAAGTCTCT 420  
 CAAAATTACA TCGCCGAGAC GCGCGACAA TCTCTCAGCG CGGCACATC CTCACCTCA 480  
 TCGCAAGCGC CTACCAATT CAATATCAGT TCGGACAT ACCAGTCCCG TATACCTTC 540  
 CTCTTACCTT CAGCTCTCT CTCTTCTCT TCTTCTCTT TCTTCTCTT TCTTCTCTT 600  
 CAGCTCTCTT CAGCTCTCT TCTTCTCTT TCTTCTCTT TCTTCTCTT TCTTCTCTT 660  
 CAGCTCTCTT CAGCTCTCT TCTTCTCTT TCTTCTCTT TCTTCTCTT TCTTCTCTT 720  
 CAGCTCTCTT CAGCTCTCT TCTTCTCTT TCTTCTCTT TCTTCTCTT TCTTCTCTT 780  
 CAGCTCTCTT CAGCTCTCT TCTTCTCTT TCTTCTCTT TCTTCTCTT TCTTCTCTT 840  
 CAGCTCTCTT CAGCTCTCT TCTTCTCTT TCTTCTCTT TCTTCTCTT TCTTCTCTT 900

3) INFORMATION FOR SEQ ID NO:4:

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGTATGAAC ACGGCCCGGT CCGATAACTT CCAGCTGTCC CAGGGTGGGC AGGGATTCCG	60
CATTCCGATC GGGCAGGCGA TGGCGATCGC GGGCCAGATC CGATCGGGTG GGGGGTCACC	120
CACCGTTCAT ATCGGGGCTA CCGCCTTCTT CGGCTTGGGT GTGTGCGACA ACAACGGCAA	180
CGGGCGACGA GTCCAACGCG TGGTCGGGAG CGCTCCGGCG GCAAGTCTCG GCATCTCCAC	240
TGGCGACGTG ATCACCAGCG TCGACGGGCG TCGGATCAAC TCGGCCACCG CGATGGCGGA	300
TGGCGTTAAG GGGCATCATC CCGGTGACGT CATCTGGTG AACTGGCAAA CCAAGTCGGG	360
TGGCACGGGT AAGGCGAAGC TCACATTGGC CGAGGGACCC CCGGCCTGAT TTGGTCGYCG	420
ATACCAGCCG CCGCCCGCCC AATTGGA	447

2 INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 604 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTCCCACTCC GGTGCGGAG CATCTCGCCC AGCAAAATGT TGGCAGCCCG CCAACGGAAT	60
CGGCTGATCC CAGGTGGAAG TTGTGGAAG CCGCCCGGGT CGAAGTATCC CTCGATCCCT	120
AAGCGGCGCA CCGCGAGCGT CGAATGCGC CGACTGAGGA AGCGGCGAAT TTGCGCGGGC	180
CGGGCCACCG NGAGCGCGCG AATGGCGCGA CTGAGGAGGT CGNCACTCAT CCCCAGNGTC	240
ATCGAATCAA CCGTATTCCT CCGTGGCGG CATTTGAGA ATTAATTAAG TGAAGCGCAA	300
CAATGATCC AAAAGGGGAG AGACGTAAT ATTTTCTGCT GTTATTAAT CATTGCTCC	360
CTGCGCGCT ATAGGATCT CATTGCTCA AACTGATCT GAGGAAAGG TTGCTGCGC	420
CAAGCGGCAAT CGGTGCGCAAT CCGTGGGCT TCGCGCAAT CAGAGAGCG TTGATGNGA	480
CAAGAGGCTG CAGCAGGCTG AATTGNGGCG CCAANAAGG CAGAGGNGG CAGAGTNGCT	540
CAAGAGGCTG AAGAGGCTG CAGAGGCTG CAGAGGCTG CAGAGGCTG CAGAGGCTG	600

(A) LENGTH: 633 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGCANGTCG AACCACTCA CTAAAGGGAA CAAAAGCTNG AGCTCCACCG CGGTGGCGGC	60
CGCTCTAGAA CTAGTGKATM YYYCKGGCTG CAGSAATYCG GYACGAGCAT TAGGACAGTC	120
TAACGGTCCT GTTACGGTGA TCGAATGACC GACGACATCC TGCTGATCGA CACCGACGAA	180
CGGGTGCGAA CGCTCACCCT CAACCGGCGG CAGTCCGCGA AGCGGCTCTC GCGCGCGCTA	240
CGGGATCGGT TTTTCGCGGY GTTGGYCGAC GCGGAGGYCG AGGAGGACAT CCACCTCCTC	300
ATCTTCACCG GYCGCGATCC GGTGTTCTCC GCGGACTGG ACCTCAAGCT AGCTGGCGCG	360
CGAGACCGCG CTGCGCGACA TCTACCGCG GTGCGCGCG ATGACCAAGC CGGTGATCGG	420
CGCGATCAAC GCGCGCGCG TCACCGCGCG GTTCGAAGTC GCGCTGTACT GCGACATCCT	480
GATCGCGTCC GAGCAGCGCG GCTTCGCGCA CACCGACCG CGGGTGGGGC TGCTGCGCAC	540
GTGGGACTC AGTGTGTGCT TCGCGCAAAA GGTGCGCATC GGNCTGGGCT GGTGGATGAG	600
CGTGACCGCG GACTAGCTGT CGGTGACCGA CGC	633

2. INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS

A. LENGTH: 1362 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TAAGTACAA TGTCTGCGA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	60
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	120
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	180
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	240
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	300
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	360
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	420
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	480
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	540
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	600
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	660
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	720
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	780
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	840
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	900
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	960
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	1020
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	1080
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	1140
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	1200
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	1260
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	1320
TAAGTACAA TCGCGCTCA TCGCGCTCA CAGCGCATC GAGCTGCGT TCGCGAGAT	1362

GCGGCGGCGAC CGTTCGGGCCC GGATGTGGGC GCCGAATACC TGGGCACCGC GGTGCAATTC 540  
 CACTTCATCG CACGCGCTGGT CCGTGGTGCTG CTGGACGAAA CCTTCCTGCC GGGGGGCCCC 600  
 GCGGCGCAAC AGCTCATGCC GCGCGCGCGT GGACTGGTGT TCGCCCGCAA GGTGCGCGCG 660  
 GAGCATCGGC CGGGCCGCTC CACCGCGCGG CTGGAGCGGC GAACGCTGCC CGACGATCTG 720  
 GCATGGGCAA CACCGTCCGA GCCCATAGCA ACCGCGTTCC CCGCGCTCAG CCACCACCTG 780  
 GACACCGCGC CGCACCTGCC GCCACCGACT CGTCAGGTGG TCAGGCGGGT CGTGGGGTCC 840  
 TGGCACGGCG AGCCAATGCC GATCAGCAGT CGCTGGACGA ACCAGCACAC CGCCGAGCTG 900  
 GCGGCGGACC TGCACGCGCC CACCGCTCTT GCGCTGCTGA CCGGCGTGGC CCGGCATCAG 960  
 GTACCGSACC AGGACGTGCC GCGGCGCGCA TCGGTGCTCG ACACCGATGC GCGCGTGGTT 1020  
 GCGGCGCTCG CCTGGGCGCG GTTCACCGCC GCGGCGCGCA TGGGCACCTG GATCGCGCGT 1080  
 GCGGCGGAGG GCCAGGTGTC GCGGCGMAAC CGACTGGGT GAGTGTGCGC GCGCTGTGCG 1140  
 TAGGCTGTCA TCGCTGGCCC GAGGGATCTC GCGGCGCGCA ACCGAGGTGG CGACACAGGT 1200  
 GGAAGCTGCG CCGACTGGGT TCGGCGCGCA CCGGCTGCTG GCGGTTGCGT TGGCGCGACT 1260  
 GCGGATCAG CTCGGGCGCG GCGGTTGCGC GAAGGTCCAG CTCACCGTGC CCGACCGCAA 1320  
 GCGGATCAG CTCACCGCGG CTCACCGTGC GCGGCGAAGG AA 1360

# 1. INFORMATION FOR SEQ ID NO: 1

## 1.1 SEQUENCE CHARACTERISTICS

- A. LENGTH: 1458 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## 1.2 SEQUENCE DESCRIPTION (Seq. ID No. 1)

GAGCAATTC GATATGCCC GCGACGCTAC GCGAATTC GCGAATTC GCGAATTC 120  
 GATATGCTTC GATGAGGAC ATTGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT 120  
 TCGATGAGGT GCGGCTGCTT TACATCATTT GCGGCGCGCG GCGGCGCGCG GCGGCGCGCG 130  
 TCGGCGCGCTT TCGGCGCGCT GCGGCGCGCT GCGGCGCGCT GCGGCGCGCT GCGGCGCGCT 240  
 GCGGCGCGCTT TCGGCGCGCT GCGGCGCGCT GCGGCGCGCT GCGGCGCGCT GCGGCGCGCT 300

CGATTGAGGA TTGCTGCAA TCGATCTTTG TGACGCTGGG ACAGGCCGCG GAGCTGCAGC 540  
 GGGCTGGAGG CGGCACCGGA TATGCGTTCA GCGACCTGCG ACCCGCCGGG GATCGGGTGG 600  
 CCTCCACGGG CGGCACGCGC AGCGGACCGG TGTCGTTTCT ACGGCTGTAT GACAGTGCCG 660  
 CGGGTGTGGT CTCCATGGGC CGTCGCGGCG GTGGCGCCTG TATGGCTGTG CTTGATGTGT 720  
 CGCACCCGGA TATGTGTGAT TTCGTCACCG CCAAGGCCGA ATCCCCCAGC GAGCTCCCCC 780  
 ATTTCAACCT ATCGGTTGGT GTGACCGAGC CGTTCCTGCG GGCCGTGCGA CGCAACGGCC 840  
 TACACCGGCT GGTCAATCCG CGAACCGGCA AGATCGTCGG GCGGATGCCC GCCGCGGAGC 900  
 TGTTCGAGCG CATCTGCAAA GCGCGCCAGC CGGCTGGCGA TCCCGGGCTG GTGTTTCTCG 960  
 AGACGATCAA TAGGSCAAA GCGGTCCCGG GGAGAGGCGG GATCGAGGCG ACCAACCCCT 1020  
 CGCGCGAGGT CCGACTGCTG CTTACGAGT CATGTAATCT CGGCTCGATC AACCTCGCCC 1080  
 GATGCTCCG CGACGCTCGG GTGACTGCG ACCGCTCGA GGAGGTGCGG GGTGTGCGG 1140  
 GCGGTTCTCT TGATGACGTC ATGATGTCA GCGGCTACCC CTTCCCGGAA CTGGGTGAGG 1200  
 GCGCGCGCGC CACCGCGAAG ATCGGGGTGG GAGTCATGGG TTGCGCGGAA GTGCTTCCCG 1260  
 GATGCGGTAT TCGGTACGAC AGTGAAGAAG CGGTGCGGTT AGCCACCGCG CTCATGCGTC 1320  
 GATACAGCA GCGCGCGCAC ACGGATCGG GAGCGCTGGC CGAAGAGCGG GCGCATTCG 1380  
 GCGGTTGAG GATACCGCG GTGCGCTGT GCGCGCGAG CGCAACGCA GAGGTGAGCT 1440  
 GCGGCTGTG GAGGCGA 1450

# INFORMATION FOR SEQ ID NO. 1:

## 1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 562 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## 2. SEQUENCE DESCRIPTION: SEQ ID NO. 1:

AGGTTTAAAT GTGCTGGAT GTGCAACCGG GTGCGCTGTG AGGTACCGAG ATTAACTGGT 57  
 GGTGACCGG GTGCGCTGTG GCGATCGCG: TGTGTTAACT GCGGATCGCG GTGCTGTG 114  
 TGTGTTAACT GTGCAACCGG AGGCTCGCT: GCGCAACCGG GAGGTGAGCT 171

CCGCGGTGCA CCGCGCGCCG GTGCTCAAGG AAGGGGACGA TTGCCCCGAT TCGACGCTGG	360
CGGTCAAAGG TTTGACCAAC GCGCGCGCAGT ACTACGTCGG CGACCAGCCG AAGTTCACCA	420
TGGTGGTCAC CAACATCGGC CTGGTGTCTT GTAAACGCGA CGTTGGGGCC GCGGTGTTGG	480
CCGCCTACGT TTA CTGCTG GACAACAAGC GGTGTGTGTC CAACCTGGAC TCGCGGCCCT	540
CGPATGAGAC GCTGGTCAAG ACGTTTTCCT CCGGTGAGCA GGTAACGACC GCGGTGACCT	600
GGACCGGGAT GGGATCGGC CCGCGCTGCC CATTGCCCCG GCGGCGGATC GGGCGGGCA	660
CGTACAATCT GGTGGTACAA CTGGGCAATC TGGCTGCTT GCGGGTTCCG TTCATCTGA	720
ATCAGCCGCG CCGCGCGCCG GGGCGGTAC CCGCTCGGG TCCAGCGCAG GCGCCTCCGC	780
CGAGTCTCC CCGCGAAGGC CGATAATTAT CGATGCTGA TGGTCGATTC CGCCAGCTCT	840
GACAACCCCT CGCTGCTTC	862

10 INFORMATION FOR SEQ ID NO:10:

(1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 622 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTGATGAGCA CCGCGAAGG GTGATGCTT TCGCTGGCTG TCGAGTGCAC CAATGACAAA	60
GAAATGCTT CCGCGAAGAT GTGGAAGTA GTGCTGGCTG TTGCTGCTTC GAACTGCTGA	120
TTGCGAAGG GTGCTGCTT GACCAAGCTT GAGGAGGCG CGATGAGCG CCGGACCGG	180
TTGCTGCTTC CGTGGCTGC CAAAGCGCG GCGGCGAGG TGGGCTTAC GTTTCAGGAT	240
TTGCTGCTTC TACGCGCA AGTGCATC GTGCTGCTTC AGTGCATC GTGATGAGG	300
TTGCTGCTTC GTGCTGAGG GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC	360
TTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC	420
TTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC	480
TTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC	540
TTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC GTGCTGCTTC	600

(A) LENGTH: 1200 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

GGCGCAGCGG TAAGCCTGTT GGCCGCCGGC AACTGGTGT TGACAGCATG CGGCGGTGGC      60
ACCAACAGCT CGTCGTGAGG CCGAGGCCGA ACGTCTGGGT CGGTGCACTG CCGCGGCAAG      120
AAGGAGCTCC ACTCCAGCGG CTCGACCCCA CAAGAAAATG CCATGGAGCA GTTCGTCTAT      180
CGCTAGCTGC GATCGTGCCG GGGCTACAGC TTGGACTACA ACGCCAACGG CTCGCGTCCG      240
CGGCTGACCC AGTTTCTCAA CAACGAAACC GATTTCCGGG GCTCGGATGT TCGGTTGAAT      300
CGCTCGACCC CTCACCTGA CCGCTCGGCG CAGCGGTCCG GTTCGGCGCG ATGGGAGCTG      360
CGGACGGTGT TCGGCCCGAT CGCGATCACC TACAATATCA AGGCGGTGAG CACGCTGAAT      420
CTTCACCGAC CCACTACCCG CAAGATTTTC AACGGCACCA TCACCTGTG GAATGATCCA      480
CAGATCCAAG CCGTCAACTC CGGCACCGAC CTGCGGCCAA CACCGATTAG CGTTATCTTC      540
CGGACCGACA AGTCCCGTAC CTCGGACAAC TTCAGAAAT ACCTCGACCG TGTATCGAAG      600
CGGCGGTGGG CCAAAGGCCG CAGCGAAACG TTCAGCGGGG CGCTCGCGCT CGGCGCCAGC      660
CGGAACAACG CAACGTCCGC CTTACTGCAG ATGAGGAGCG CGTCCATCAC CTACAACGAG      720
CGCTGCTTTG CGCTCGGTAA CGAGTTGAAT ATGAGGAGCA TCATCACCTC CGGCGGTCCG      780
GATTAATGTC CGATCAGCAC CGAGTGGGTG GTTAAGACAA TCGCGCGCGG CAAGATCATG      840
CGAATAGGCA ACGACCTCGT ATTCGACAGC TCGTCTTTTC ACAGACCCAC CGAGCTTGGC      900
CTTTACAGCA TCGTCTCTGC GAGCTATGAG ATCTCTGCTT GAAATACCC CGATCCGAGC      960
CGCTGCTTTC GATTAAGTTC GTTATGTA ACCTGATTC TATTAAGCA GAAAGCTATC      1020
CGCTGATTC ATTCATTCG GTTATGTA ACCTGATTC TATTAAGCA GAAAGCTATC      1080
GATTAATTC ATTCATTCG GTTATGTA ACCTGATTC TATTAAGCA GAAAGCTATC      1140
CGCTGATTC ATTCATTCG GTTATGTA ACCTGATTC TATTAAGCA GAAAGCTATC      1200
  
```

SEQUENCE INFORMATION FOR SEQ ID NO: 11

SEQUENCE INFORMATION FOR SEQ ID NO: 11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCAAGCAGCT GCAGGTCGTG CTGTTGACG AACTGGGCAAT GCCGAAGACC AAACGCACCA	50
AGACCGGCTA CACCACGGAT GCCGACGGCC TGCAGTCGTT GTTCGACAAG ACCGGGCATC	120
CGTTTCTGCA ACATCTGCTC GCCCACCAGC ACGTCACCCG GCTCAAGGTC ACCGTCGACG	180
GOTTGCTCCA AGCGGTGGCC GCCGACGGCC GCATCCACAC CACGTTCAAC CAGACGATCG	240
CCCCGACCGG CCGGCTCTCC TCGACCGAAC CCAACCTGCA GAACATCCCG ATCCGCACCG	300
ACCCCGCCCG CCGGATCCCG GACCGCTTCC TCGTCGGGGA CCGTTACGGC GAGTTGATGA	360
CGCCCGACTA CAGCCAGATC CAGATCGCGA TCATCGGGCA CCGTCCCGGG GACGAGGGCC	420
TCATCGAGGC GTTCAACACC GCGGAGGACC TGTATTGCTT CCGCGCTCC CCGGTGTTCC	480
GTGTCCCAT CAGCAGGTC ACCGCGAGT TCGCGCCCG CCGCAAGGCC ATGTCTACG	540
GGTGTGTTTA CCGGTTGAGC GCCTAGGGCC TCTCCAGCA GTTGAAAATC TCCACCGAGC	600
AAGTCAACGA GCAGATGGAC GCGTATTTCG CCGGATTCGG CCGGGTCCGC GACTACCTGC	660
CGCGCTACT CAGCCCGGCC CCGAAGGACC GCTACACCTC CACGGTGGTG GCGCGTCCGC	720
GCTACCTGTC CAGGCTGGAC AGCAGCAACC GTCAAGTCCG GAGGCCCCCG GAGCGGGCCG	780
CGCTGAACCT CCGATCCAG CCGAGCCCGC TCGACATCAT CAAAGTGGCC ATGATCCAGC	840
TGGAAAGCT CCGAAGCA CCGAGCTCC GCTGCGCAT GTGCTCCAG GTCCAGCAGC	900
AGCTGCTGTC TGAATGCGC TCGGTCGAA CCGAGCTGCT CGAGGCGCTC GTCCCGACA	960
AGATGCGCGC TCGTACCGC CTCAGCTCC GCTGCGCAT GTGCTGCGC TACCGCGCA	1020
GCTGCGCAT CCGGCGCA TCGTCCCGA GCTGCGCAT CCGGCGGAA TCGCGCAT	1080
CGTTCGAT CAGTTCGAT CAGTTCGAT CAGTTCGAT CAGTTCGAT CAGTTCGAT	1140
CAGTTCGAT CAGTTCGAT	1200

SEQUENCE INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS

- A LENGTH 1201 base pairs
- E TYPE nucleic acid
- C STRANDEDNESS single
- D TOPOLOGY linear

TGGGGGCTCG GGTTCGGCAT GGTCAAACAG GTGGTGTCTA ACCACGGCGG ATTGCTGCGC 120  
ATCGAAGACA CCGACCCAGG CGGCCAGCCC CCGGAACGT CGATTTACGT GGTGCTCCCC 180  
GGCCGTGGA TGGCGATTCC GCAGCTTCCC GGTGCGACGG CTGGCGCTCG GAGCACGGAC 240  
ATCGAGAACT CTCGGGGTTC GGCGAACGTT ATCTCAGTGG AATCTCAGTC CACGCGCGCA 300  
ACCIAGTGTG GCAGTTACTG TTGAAAGCCA CACCCATGCC AGTCCACGCA TGGCCAAGTT 360  
GGCCCGAGTA GTGGGCCTAG TACAGGAAGA GCAACCTAGC GACATGACGA ATCACCACG 420  
GTATTGGCCA CCGCCGACGC AGCCGGGAAC CCCAGGTTAT GCTCAGGGGC AGCAGCAAAC 480  
GTACAGCCAG CAGTTGCACT GCGGTTACCC ACCGTGCGCG CCGCCGACGC CAACCCAGTA 540  
CGGTCAACCC TACGAGGCGT TGGGTGCTAC CGCCCGCGCT CTGATACCTG GGTGATTCC 600  
GACCATGACC CCGCCTGCTG GATGCTTCG CCAACGCGCT GGTGACGCA TGTTCGCGAT 660  
CGCCCGGCTG ACATAGCGCG TGGTGTCCCG CGGCATGCGC GCGCGGCGCG CATGCGTGGT 720  
CGGGTTCAAC CCGGCACCGG CCGGCCCGAG CGCGCGCGCA GTGGCTGCGA GCGCGCGCGC 780  
AAGCATCCCC GCAGCAAACA TCGCGCGGGG GTCGGTGGA CAGGTGCGCG CCAAGGTGGT 840  
CGCCAGTGTG GTGATGTTGG AAACGATCT GCGCGCGCG TCGGAGGAGG GTCGCGGAT 900  
CATTCGTGT GCGGAGGGGC TGATCTTGAC CAACAACCA GTGATGCGCG CCGCGCGCAA 960  
CGCTGCGCTG CGCAGTCCCG CGCGCAAAAC GAGGTAAAC TTCTGTGAGG CGCGGACGCG 1020  
ACGCTTGACG GTGCTGGGGG CTGACCGCAC TATGATATG GCGCTGCTCG GTGTTGAGG 1080  
GTGCTGCGCG CTGACCGCGA TGTGCGTGGG TTGCTGCTCG GAGGTGAGG TCGGTGAGCG 1140  
GTGCTGCGCG ATCGGCTCGC CGCTCGGTTT GAGGCGCACT GTGACCGCGG GTGCTGCGAG 1200  
GTGCTGCAAC GGTGAGTGT GAGGACCGCG GAGGCGCGG AAGGAGAAAG GGTGCTGCGA 1260  
GTGCTGCAAC GGTGAGCGCG GTGACCGCG GTGAGTGTG GTGCTGCGG GTGCTGCGAG 1320  
GTGCTGCAAC GTGCTGCGAG GTGCTGCGG GTGCTGCGG GTGCTGCGG GTGCTGCGAG 1380  
GTGCTGCAAC GTGCTGCGAG GTGCTGCGG GTGCTGCGG GTGCTGCGG GTGCTGCGAG 1440  
GTGCTGCAAC GTGCTGCGAG GTGCTGCGG GTGCTGCGG GTGCTGCGG GTGCTGCGAG 1500  
GTGCTGCAAC GTGCTGCGAG GTGCTGCGG GTGCTGCGG GTGCTGCGG GTGCTGCGAG 1560  
GTGCTGCAAC GTGCTGCGAG GTGCTGCGG GTGCTGCGG GTGCTGCGG GTGCTGCGAG 1620

GTGATGAAGG TCGCCGCGCA GTGTTCAAAG C

1571

(12) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCCACCCCG GTGCGCGCCG GTGTAGAAGT AGTGGATCCC CCGGGCTGCA GGAATTGCGC	60
ACGAGGATCC GACGTCCAG GTTGCGAAC CCGCGCCCGG GGAAGTATCG GTCCATGCGT	120
AGCCCCCGCA CCGCAGCCG CGAATGCGG TGAGTGAGCA GCGCGGCAAT TTGCGCGCGC	180
CGCGCGACCG CGAGCCCCCG AATGGCGCGA GTGAGGAGCG GCGCAGTCAT GCGCAGCGTG	240
ATCCAATCAA CCTGCATTCC GCGTCCCGCG CCATTTCACA ATCGAGGTAG TGAGCGCAAA	300
TGAATGATGG AAAACGGCGG GTGACGTCCG GTTTTGTGGT GGTGCTAGGT GCGTGCCTGG	360
GCTTGTGGGT ATCAGGATGT TCTTCCCGCA AACGTGATGC CGAGGAACAG GGTGTTCCCG	420
TGAGCCCCGAC GCGGTCCGAC CCGCGCGCTC TCGCGGAGAT CAGGCAGTCC GTTGATCGCA	480
CAAAAGGCTT GACGAGCGTG CAGGTAGCGG TCGGAACGAC CGGAAAAGTC CACAGCTTGC	540
TGGTATTAG CAGTCCGAT CTCGACGTCC CGGCAATGT GGTGCGCGCA AAGGCGGTAT	600
CGAGGTACAA CGAGGAGCAG GTGTGCGCT TTGCGGTA TA AGGCGACAG ATGTGCGTGA	660
CACTGTTGCA CGACTGGAGC AATGTGCGGT CGATTTCGCA ACTGTCAACT TCAGCGCTGC	720
TGGATGCTGC GCGTGGGTTG ACGCAGCTGC TGTCCGCTGT TACGAACTTC CAGCGGCAAG	780
TTGTTAAAT ATAGAGTGA GTTGAGCTT TAAAGATTA TCGAATATC TCGGAGCTT	840
TTTAAATAT TTTGATTTT TTTTAAAGA TTTAAAGTA TTTAATTT TCGTTTCTT	900
TTTAAATTT TCGAGCTTT GTGTAGGCA TGTGAGCTT TGTATCGAA TCGATTTCG	960
TGTGAGCTT GAAATGGAAG GAAATGCTTA AGGTGAGTA GCGGAAAGTT GGTGAGCTT	1020
TTTAAATGAA AGGCGCTTTT GAAATGCTTT AAGGCAAG	1058

END OF SEQUENCE FOR SEQ ID NO:14

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAACAGGC	60
GGCGGCGGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT	120
CATTGAGCAG GACATGGCCG TGGACAGCGC CGGCAAGATC ACCTACCGCA TCAAGCTCGA	180
AGTGTCTTTC AAGATGAGGC CGGCGCAACC GCGCTAGCAC GGGCCGGCGA GCAAGACGCA	240
AAATCGCAGC GTTTGCGGTT GATTCGTGCG ATTTTGTGTC TGCTCGCGGA GGCCTACCAC	300
CGCGGGCCCA GGTGCGCGTG CTGCCGTATC CAGGCGTGCA TCGCGATTCC GCGGGCCACG	360
CGCGAGTTAA TGCTTCGCGT CGACCCGAAC TGGCGGATCC GCGCGGAGC TCATCGATGA	420
CGGTGCGCAG CGCGTCGATC CGCGAGTTCC CGAGGAAAC GTGCTGCGAG CGCGGTAGGA	480
AGCGTCCGTA GCGCGCGGTC CTGACCGGCT CTGCGTCCGC CCTCAGTGCG GCGAGCGAGC	540
GC	542

(D) INFORMATION FOR SEQ ID NO:15:

## 1) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 918 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:16

AGTTGCGGCG CGCGGCTTCC TTGCGCGCAT TGCGGCGGTC CGCGATCAGC TGCGCATCGC	60
AGCGATCAGC CGCGGCTTCC CGCGGCGGTC CGCGGCGGTC CGCGGCGGTC ATGCGACCGC	120
ATGCGGCTTCC CGCGGCGGTC CGCGATCAGC ATGCGGCTTCC CGCGGCGGTC ATGCGGCTTCC	180
ATGCGGCTTCC CGCGGCGGTC CGCGGCTTCC ATGCGGCTTCC CGCGGCGGTC ATGCGGCTTCC	240
ATGCGGCTTCC CGCGGCGGTC CGCGGCTTCC ATGCGGCTTCC CGCGGCGGTC ATGCGGCTTCC	300
ATGCGGCTTCC CGCGGCGGTC CGCGGCTTCC ATGCGGCTTCC CGCGGCGGTC ATGCGGCTTCC	360
ATGCGGCTTCC CGCGGCTTCC CGCGGCTTCC ATGCGGCTTCC CGCGGCGGTC ATGCGGCTTCC	420
ATGCGGCTTCC CGCGGCTTCC CGCGGCTTCC ATGCGGCTTCC CGCGGCGGTC ATGCGGCTTCC	480

CGGCGCGCGGCG GGACCCACCG GTCCCGCGCGA TCCCGCGGTT GCCGCGGGTG CCGCGCGCAT	720
TGGTGGCTCT GAAGCCGTTA GCGCGGGTTC CGCGGGTTCC GCGGGTGGCG CCGTGGCGCG	780
CGGCGCGCGGCG GTTGCCGTAC AGCCACCCCG CGGTGGCGCG GTTGCCGCCA TTGCGCGCAT	840
TGCGCGCGGT GCCCGCATTC CCGCGGTTC CGCGCGCACG CCGGGNTTGG CCGCGCGCGCG	900
CGCGCGCGCGCG CGC	913

## 2. INFORMATION FOR SEQ ID NO:17:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1872 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## X1. SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTACGTTG GTGTAGAAA ATCTGCGCG CCGGACCGTT AAGGCTGGGA CAATTTCTGA	60
TAGCTACCGG GACACAGGAG GTTACGGGAT GAGCAATTCC CCGCGCGCGT CACTCAGGTG	120
GTGATGTTTG GTGAGCGTGC TGGCTGCGGT CGGCTGCGG GTGGCCACCG CCGCGCGCGCA	180
CGCGCGCGCG CCGCGCGTGT CCGAGGACCG GTTCGCGGAC TTCCCGCGCG TGCGCGCTCGA	240
CGCGTTCGCG ATGGTCCCGG AAGTGGCGCG ACAGGTGGTG AACATCAACA CCAAACTCGG	300
TTAGACAAAC CGCGTGGCGG CGCGGACCGG CATCGTCATC CATCGCAACG GTGCTGCTGT	360
TAGTACAAAC GAGTGTATCG CGCGCGCGAC TGACATCAAT CGTTGAGCG TCGGCTGCGG	420
TGAAAGCTAC CGCGTGGATG TGCTGCGGTA TGACCGCGCG GAGGATGTGG CGGTGCTGCA	480
TGTCGCGCGT CGCGGTGGCG TGCTGCGGCG CGCATCGGT CGCGCGCGTG TGCTTGGTGA	540
CGCTTCTTTT CGATCGCGCA AAGCTTTTGA TACGCGCGCA AAGCTTTTGA GTTCTGTTTG	600
CGCTTCTTTT CGCTGCGGTA AAGCTTTTGA CGCTGCGGAT TACGCGCGCA GTTCTGTTTG	660
TACGCGCGCA CGCTGCGGAT AAGCTTTTGA TACGCGCGCA CGCTGCGGAT TACGCGCGCA	720
TGCTTCTTTT CGCTGCGGAT AAGCTTTTGA TACGCGCGCA CGCTGCGGAT TACGCGCGCA	780
TGCTTCTTTT CGCTGCGGAT AAGCTTTTGA TACGCGCGCA CGCTGCGGAT TACGCGCGCA	840
TGCTTCTTTT CGCTGCGGAT AAGCTTTTGA TACGCGCGCA CGCTGCGGAT TACGCGCGCA	900

CATCAACTCG GCCACCCGCA TGGCCGACCG GCTTAACCGG CATCATCCCC GTGACGTCAT 1080  
 CTCGGTGAAC TGGCAAACCA AGTCGGGGCG CACGCGTACA GGGAACGTGA CATTGGCCGA 1140  
 GGGACCCCCG GCCTGATTTC TCGCGGATAC CACCCGCGCG CCGGCCAATT GGATTGGCGC 1200  
 CAGCCGTGAT TGCCGCGTGA GCGCCCGAGT TCGGTCTCCC GTGCGCGTGG CATTGTGGAA 1260  
 GCAATGAACG AGGLAGAACA CAGCCTTGAG CACCCTCCCC TGCAGGGCAG TTACCTCGAA 1320  
 GCGCGTGTGG TCGAGCATCC GGATGCCAAG GACTTCGGCA GCGCGCGCGC CTTGCCCGCC 1380  
 GATCCGACCT GGTTTAAGCA CGCCGTCTTC TACGAGGTGC TGSTCCGGGC GTTCTTCGAC 1440  
 GCGAGCGCGG ACGGTTCCGN CGATCTCCCT GGAATCATCG ATCGGCTCGA CTACCTGCAG 1500  
 TGGCTTGGCA TGAATTCAT GTGTTGCGG GGTTCCTAGC ACTCACCCTT GCGCGACCGC 1560  
 GGTACGACA TTGCGGACTT CTACAAGGTG GTGCCCCAAT TCGGCACCGT CGAGGATTCG 1620  
 GTGCGCGTGG TCGACACCCG TCACCGGCGA GGTATCCGCA TCATCACCGA CTTGGTGATG 1680  
 AATCACACCT CGGAGTCGCA CCCCTGGTTT CAGGAGTCCC GCGCGGACCG AGACGGACCG 1740  
 TACGGTGAAT ATTACGTGTG GAGCGACACG AGCGAGCGCT ACACCGACCG CCGGATCATC 1800  
 TTGCTGACCA CGGAAGAGTC GAACTGGTCA TTGATGCTG TCGCGCGACA GTTNGTACTG 1860  
 CCACCGATTC TT 1920

# 2. INFORMATION FOR SEQ ID NO.13

## 1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 1482 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## 2. SEQUENCE DESCRIPTION: SEQ ID NO.13

ATGAGCAAA AGCTGATTCG GAGGAACGCG GTGTTGCGCT GAGGCGAAT AGTTCGCAAT  
 ATGCTGCTCT AGCGGAGATC AGCGAGTCCG TCGATCGGAT AAAAGGCTTC AGCAGGCTTC  
 AGTATGCGCT CGGAACAACG TGGAAAGTCC ACAGGTTGCT GGGTATTACT AGTCCCGATC  
 TGAATGTCGG GCGCAATCCG GTGCTGCGAA AGCGGCTATG CATTTACAAC GACGAGCAGG  
 GTGCTGCTCT TCGGCTACAA TCGATCAACA TCGATGCAAA AGTCTGCAAA GACTGAGCA  
 GAGTATGCTC AGTTCGCTCT GTGCTGCTCT GTGCTGCTCT GTGCTGCTCT GTGCTGCTCT

GCGCCCAAGAG TGCAAGGCCG GCGACCGTGT GGATTGCCCA GGACGGCTCG CACCACCTCG 540  
 TCCGAGCGAG CATCGACCTC GGATCCGGGT CGATTCAGCT CACSCAGTCG AAATGGAACG 500  
 AACCCTGTCAA CGTCGACTAG GCCGAAGTTG CGTCGACCGG TTGCTCGAAA CGCCCTTGTG 660  
 AACGGTGTCA ACGGCACCCG AAAACTGACC CCCTGACGGC ATCTGAAAAT TGACCCCTTA 720  
 GACCGGGCGG TTGGTGGTTA TTCTTCGGTG GTTCCGGCTG GTGGGACCGG GCCGAGCTCG 780  
 CGGTCTTTGA GCGCGTAGCT GTCCCTTTG AGGGCGACGA CTTCAGCATG GTGGACGAGG 840  
 CGGTGATCA TGGCGGCAGC AACGACGTG TCGCCGCCGA AAACCTCGCG CCACCGCGCG 900  
 AAGGCTTTAT TGGACGTGAC GATCAAGCTG GCGCGCTCAT ACCCGGAGGA CACCAGCTGG 960  
 AAGAAGAGGT TGGCGGCTC GGGCTCAAA GGAATGTAAC CGACTTCGTC AACGACCAGG 1020  
 AGCGGATAGC GGCCTAACCG GGTGAGTTG CGGTAGATGC GCGCGGGCTG GTGAGCTCG 1080  
 GCGAACCCTG CTACCCATTC GCGCGGGTG GCGAACGCA CCCGATGACC GGTCTGACAC 1140  
 GCGCGTATCG CCAGGCGGAC CGCAAGATGA GTCTTCGCGG TCCAGGCGG GCGCCAAAAA 1200  
 CACGACGTTA TCGCGGGCGG TGATGAAATC CAGGGTGCGG AGATGTGCGA TGGTGTGCGG 1260  
 TTGAGGCGCA CGAGCATGCT CAAAGTCGAA GTCTTCGAA CACTTCGAA CCGGGAAGCG 1320  
 GCGCGCGCGG ATGCGCGGCT CAGGACCATG CGACTCGCGG GCTGACACTT CCGGCTGAG 1380  
 GCAAGCGCGG AGGATATGTT GTGGGTGCA GTCTCGCGG TCGCGCGCAT CCGCGAGCG 1440  
 TGACGTCAG TCGCGAGCG TCGAGCTTT TAATGCTTT TT 1480

## 2. INFORMATION FOR SEQ ID NO.19

### 1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 1489 base pairs
- B. TYPE: UNKNOW
- C. STRANDEDNESS: SINGLE
- D. TOPOLOGY: LINEAR

### 2. SEQUENCE DESCRIPTION: SEQ ID NO.19

GAATTGCTTA CAGCGCGCGG ATAGCTTCTT GCGCGCGCGG TACCAJATTC CTGAGGGTT 80  
 GTCTTCTTGG GCGCGCGCGG GTGCGAGCA CTGACCGGT CAGGAGCTTC AACGCGCTCA 120  
 TCTGATCTTA GTCTCTCTTC ATGCGCGGTA GTCTCTCTTC ATGCGCGCTTC 160

GGAGAACTTC GATCCCGAGG GCGTCTCTGGG GGGTATCTAC CGTATCACC CGGCCACCGA 360  
 GCAACGCACC AACAAGGNGC AGATCCTGCC CTCGGGGGTA GCGATGCCCG CGGCGCTGCC 420  
 GGCAGCACAG ATGCTGGCCG CCGAGTGGCA TGTCCGCCCC GACGTGTGGT CCGTGACCAG 480  
 TTGGGGCGGAG CTAAACCGCG ACGGGGTGGT CATCGACACC GAGAAGCTCC GCCACCCCGA 540  
 TCGGCCCGCG GCGGTGCCCT ACGTGACGAG AGCGCTGGAG AATGCTCGGG GCCCGGTGAT 600  
 CCGGCTGTCT GACTGGATGC GCGCGGTCCC CGAGCAGATC CGACCGTGGG TGCCGGGCAC 660  
 ATACCTCACC TTGGGCACCG ACGGGTTCGG TTTTTCGGAC ACTCGGCCCG CCGGTCTCTG 720  
 TTACTTCAAC ACCGACGCGG AATCCGAGT TGGTCCGGGT TTTGGGAGGG GTTGGCCGGG 780  
 TCGACGGGTG AATATCGACC CATTCGGTGG CCGTCTGGG CCGGCCCGCC AATTACGCGG 840  
 AATTGACGAA CCGTGGGGGT TCGGCCCGAN TAAGTT 876

# 2. INFORMATION FOR SEQ ID NO:20:

1. SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1021 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

# X1. SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATCCGCGCGG CTCGACGAA TTCGCGACCA GAGACAAAT TCGACGCTT AATGACGAA 60  
 GAGATTGATA ACCAATTGAC AGCGGCACAA CAATATGTCC CGATCGCGGT TTATTTCGAC 120  
 ACCBAAGACC TCGTCGAGTT GCGCAAGCAT TTTTACAGCT AAGCGGTGCA GGAACGAAAC 180  
 CATGCAATGA TCGTCTGCA ACACCTCTCT GACCGCGACT TTCTGTGCA AATTCGCGCC 240  
 CTAGACACCG TCGGAAACCA TTTCGACAGA TCGCGCGAGG CATTCGCGCT GCGGCTCGAT 300  
 AGGAAACGCA TACTTAACA TAAGTAAAT TCGTCAAGT TCGTGGCGG GACGAGCGCT 360  
 TATTTCTCTG CGCAATATG TAAGTAAAT TATTCGAG AAACATGTA AGAGGTGCTT 420  
 TTATGTCGAA CCGTGGTCTG GTTTCGAT TCGCGCGCGG CGAAGCTGTT GAGCTAGAC 480  
 AATTGCTCTG CAGCTGAAT CGATTGCGG TCGCGCGCGG CAGGCGCGCG GCAAGCTGCT 540  
 TCGCGCGCGG TGTAGATCG TCGTGGAT TAAGTAATG TCGGCTTGGT CCGTCTCTCT 600  
 TCGAGCGAGG CTTGCTCTG TCGTGGAT TAAGTAATG TCGGCTTGGT CCGTCTCTCT

GCGGTTGGCC CGACCCCGCT GCGCGCACTG CTGGTCAGGT ATCGGGGGGT CTTGGCGAGC 340  
 AACAACTGCG GCAGGAGGGG TGGAGCCCGG CGGATCCGCA GACCGGGGGG CGGAAAACGA 900  
 CATCAACACC GCACGGGATC GATCTCGGGA GGGGGGTGCG GGAATACCGA ACCGGGTGTAG 960  
 GAGCGCCAGC AGTTGTTTTT CCACCAGCGA AGCGTTTTCG GGTCAACGGN GGCNNTTAAG 1020  
 T 1021

# 2) INFORMATION FOR SEQ ID NO:21:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 321 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## 2. SEQUENCE DESCRIPTION: SEQ ID NO:21:

CTTCCCGACG AACCGAAGAA CACAACCATC AAGATGCTGA AATGATCGG CCGAGCTCTG 60  
 ACCGCGCGCG CTGCAATCGG CGCCGCTGCG GCGGCTGTGA CTTGATCAT GCGTGGCGGN 120  
 GCGGTCTGTAT ACCAGATGCA GCGGCTCTG TCGGCGCGCG CACTGCGCTT GGACCGCGNA 180  
 TCGGCGCGCTG AATGCGCGAG CGCGCGCGCA TCGACCGAGC TCGTAAACAG CTCGCGCGAT 240  
 TCGAAGCTGT CCGTTGCGAA CAAGGCGAGT CTCGCTCGAG GCGGATCGG GCGGAGCGAG 300  
 GCGGAGCGAT GCGGAGCGA 321

# INFORMATION FOR SEQ ID NO:22:

1. SEQUENCE CHARACTERISTICS:
- A. LENGTH: 173 base pairs
  - B. TYPE: nucleic acid
  - C. STRANDEDNESS: single
  - D. TOPOLOGY: linear

## 2. SEQUENCE DESCRIPTION: SEQ ID NO:22:

CATTATGCTT TCGGTTGCG GACCGCTTTT GCGGAGCTT CTTAAGCGG CTTGATCAT 60  
 GATGAGCGG GCGGAGAGG CTGAGCTCGG ATATGCTGCG GCGGTTGAG CTTGAGCGG 120  
 CTTGCTGCTT GATGCGGAA GCGGCGAGG AGCGCTTGA GACCGGATC AAGCGGATG 180  
 ACGGATGAG CCGGATGCG GCGGCTGAG CTTGATCAT CTTGCGGAG CTTGAGCGG 240

CTTACCATCG CCG

373

## 2. INFORMATION FOR SEQ ID NO:23:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 352 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

GTGACSCCGT GATGGGATTC CTGGGCGGGG CCGGTCCGCT GCCGGTGGTG GATCAGCAAC      50
TGGTTACGGG GGTGCGCCAA GGCTGGTGGT TTGCTCAGGC AGCCGCTGTG CCGGTGGTGT      100
TCTTGAAGGG CTGGTACGGG TTGCGCGATT TAGGCGAGAT CAGGCGGGGC GAATCGGTGC      150
TGATCCATCG CGGTACCGGG GTGTGCGGCA TGGCGGCTGT GCAGCTGGCT GCCCAGTGGG      200
CGGTGGAGGT TTTGCTCAGC GCCAGCCGTG GNAAGTGGGA CACGCTCCGC GCCATNGNGT      250
TTGACGACGA NCCATATCGG NGATTCCGNC ACATNCGAAG TTCCGANGGA GA              352

```

## 2. INFORMATION FOR SEQ ID NO:24:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 336 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```

AAATCGCGG TGTATTGGT TGAAGAGCG GTGCGCGATA ATGACGAA TGTATAGCG      50
AGGCTTGGCG CGGCTCATCG GTGACAGCGA TTAATCAGCA AGTTCTCTCG TATATCGCAC      100
GTAGGCTGCA TTGCTTGGC AGATCGGTTT GTAGCGTCA TCGATGTA TGTTCGCGCT      150
TGGCAAGCT CATGCTGCG GTTCTGATTA TCGGAGCGG TGTGCGGCT GTGCGGCTCG      200
TGTGAGCTA TCGAGCGAA GTGCGGCTA TGTGAGCTA TGTGAGCTA TGTGAGCTA      250
TTTTCGAGCG CGCATCGCGG TCGAAGTGG AGCGCTACAC GTGCGATGAT GATTCGAGCG      300
TGAAGAGCGA CGGCGCGAG CAGAGTGGG GTAGCGCGCG AGCGATGCTT GAGGCTCGCG      350
TGTTCGAGCA TGTGCTGCT GTGCTGCTG GTGCTGCTG GTGCTGCTG GTGCTGCTG      400
TGTTCGAGCA TGTGCTGCT GTGCTGCTG GTGCTGCTG GTGCTGCTG GTGCTGCTG      450

```

GGCGACAGCG COTCCACCAT CGACATCGAC AAGGTTGTTA CCGGCACACC CGTTCGGCGG 720  
ATCGTG 726

(2) INFORMATION FOR SEQ ID NO:25:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 580 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCGACGACG ACCAAGCTCG GCGCCACAC CCGCTATGCG TTGATGACAG CACCCGGGAT 60  
CTCGCGCGAC CATATGCAAG CATCTGGGT GCGCACTGAG CGACCTTTTG ACCGCGCGCG 120  
CTCGCGCGATG GCGCGCGCGT GAATGATTG CCGCGGGGCT TGTGCACCTG ATGAACCCGA 180  
ATAGGGAACA ATAGGCGCGT GATTGCGAG TTCAATGTC GGTATGGGTG GAAATCCAAT 240  
GCGCGGGGAT GCTCGGCGCG GACGAGGCTC GCGCAGGCGG GCGAGCCGGA ATCTGGAGGG 300  
AGCACTCAAT GCGCGCGATG AAGCGCGCGA CCGCGGACGG TCCTTTGGAA GCAACTAAGG 360  
ACCGCGCGCG CATTTGATG CGACTACAC TTGAGGCTG CCGTCGCTG GTCTGAGG 420  
TGACACCGGA CGAAGCGCG GCACTGGGTG ACGAACGAA AGGCGTTACT AGCTAAGACC 480  
AGCGCAAGG CGAATGCTG GCTTACCG CACACCTTCG GTAGATGTG AGTGTCTG 540  
TGGCGATCT ATGCGGAGGA GAATCTTGG ATACGCGCT 600

(2) INFORMATION FOR SEQ ID NO:26:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 160 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:26:

AAATGAGGCG GCGCGCGTTT TCGCGGGGCG GCGCGGCTG GCGCGAAGCG GCGCGCGCG 60  
GCGCGGCGG GTTCTTTGCG TGTGCGCGG GCGCGTGGG GCGCGGCGAA GCGCATGCT 120  
GCGCGGCGG GCGCGGCGG GCGCGGCGG GCGCGGCGG GCGCGGCGG 180

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTCGACA CGCTCGAGGC GTTCACGATC	60
CAGCGCACAC CCGACGGCGT GACCAACGGC GATGCGGCC CGTTCGCGGA GGCGGCTGCC	120
AAGGCGATGG GAATCGACAA GCTGCGGGTA ATTGATACCG GAATGGACCC CGTCGTCGCT	180
GAACGCGAAC AGTGGGACGA CCGCAACAAC ACCTTGGCGT TGGCGCCCGG TGTCGTTGTC	240
GCCTACGAGC GCAACGTACA GACCAACGCC CG	272

2. INFORMATION FOR SEQ ID NO:28:

1. SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 317 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCAGCCGGTG GTTCTCGGAC TATCTGCCCA CGGTGACGCA GCGCGAGCTG CCGAGCTGCA	60
AGCGCATCGA GCAGACGGAT CCGTTCGCGG GTTTCATGCG GTACCTGGCG GGTATCAGCG	120
CCAGCGAGCT GAACCTGGCT GAACCGGGCT GGTTCATGCG GTTCGACGCG GCGACGATCC	180
TTTCCGATCT CCGCTGCTTC GAGAGCTGCT ATCTGGTACA TCGCTGCTTC GGTGCTGCGC	240
GAATCTTCAC CCGAAGATC AAGAAAGGCT TAAAGATCGA GTCTGTGAGT ATGCTGCTTC	300
CGCTGCTGCT CCGCGCGG	317

3. INFORMATION FOR SEQ ID NO:29:

1. SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 317 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO:29:

GAATGATGAT GTATGATGAT GTATGATGAT GTATGATGAT GTATGATGAT GTATGATGAT  
GTATGATGAT GTATGATGAT GTATGATGAT GTATGATGAT GTATGATGAT GTATGATGAT

## 2. INFORMATION FOR SEQ ID NO:30:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 308 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```

GATCCCGAAG TTTGGTGAGC AGGTGGTCGA CGCGAAAGTC TGGGCGCCTG CGAAGCGGGT      60
CGGCGTTGAC GAGGCGAAGA CACGCGTGTG CGAGCTGCTG CGGCTCGTCT ACGGCGGGCA      120
TAGGTTGAGA TTGCGCGCGG CGGCGAGCGG GTAGCAAAGC TTGTGCGGCT GCATCGTCAT      180
GAGACTCGGC GATTAGGCAT TGACCATGGC GTGTACCGCG TCGCGGACGA TTTGACGCT      240
CGGTTGTGAG ACGACGTGCT CGAACGCTTT CACCGGTGAA GCGCTACCTC ATCGACACCC      300
ACGTTTGG

```

## 2. INFORMATION FOR SEQ ID NO:31:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 157 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```

CGATGACGA GCAACTGAGG TGATGATGG TCGGAGGAG GATTGAGGAC CGAGAGATC      60
TGGCGGAGG TCGCGCGCGG TAACTGATCA TAGTGACCGG CGGTAGAGGG CTGCGCGGAT      120
CGATGACGA TATTGTGCTG TGTGCTGCGG CGGTAAAGAG CGGTAAAAGA ATGTGACCGG      180
CGATGACGA TATGACAG TATGATGAG TATGATGAT CGGAGCTG TGTGACGAT      240
CGATGACGA TATGAGGG TATGAGG

```

## 2. INFORMATION FOR SEQ ID NO:32:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1533 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

GCGAGACCAT GCGCGCCCTG GACTGGTTCT AAGTACAGTC AATTCGAGGC CACCTGGTCT 180  
 ACAGAGCCGT GCGGCACTTC CAGGTGACTA TGAAAGTCGG CTTCGGCTGG AGGATTCCTG 240  
 AACCTTCAAG GCGGCGCGAT AACTGAGGTG CATCATTAAG CGACTTTTCC AGAACATCCT 300  
 GACGCGCTCG AAACGCGGTT CAGCCGACGG TGGCTCCGCC GAGGCGCTGC CTCACAAATC 360  
 CCTGCGACAA TTCGTGGGCG GCGCCTACAA GGAAGTCGGT GCTGAATTCT TCGGGTATCT 420  
 GGTGACCTG TGTGGGCTGC AGCCGGAAGA AGCGGTGCTC GACGTGGGCT GCGGCTCGGG 480  
 GCGGATGGCG TTGCGGCTCA CCGGTATCTT GAACAGCGAG GGACGCTACG CCGGCTTCGA 540  
 CATCTGCGAG AAAGGCTATG CGTGCTGCTA GGAGCACATC ACCTCGGCGC ACCCCAACTT 600  
 CCACTTCGAG CTCCTGACCA TCTACCACTC CTTCTACCAAC CCGAAAGGGA AATACCGCTC 660  
 ACTAGACTTT GCGTTTCCAT ATCGGATGC CTCGTTGAT GTGGTCTTTC TTACCTGGCT 720  
 GTTCAGCGAC ATGTTTCCGC CGGAGCTGGA GCACTATCTC GACGAGATCT CCGCGCTGCT 780  
 GAAGGCGCGC GGACGATGCC TGTGACCTA CTTCTTGCTC AATGACGAGT CGTTAGGCCA 840  
 CATCGCGGAA GGAAAGAGTG CGCACAACCT CCAGCATGAG GGACCGGGTT ATCGGACAA 900  
 CCACAAGAAAG GCGCCCGAAG AAGCAATGG CTTGCGCGAG ACCTTCGCTC GGGATGTCTA 960  
 CGGCAAGTTC GCGCTCGCGT TCGAAGAAC ATTGCACTAC GCGTCATGGA CTGGTCGGGA 1020  
 ACCAGGCTTA AGCTTCGAGG ACATGCTCAT GCGACCAAA ACCTCGAGCT AGGTTCGGAT 1080  
 CGGGAAGCA TCGGCAACT CTGGCGGGA GCGCGCTCT CCGAGGCTC ATTAGCGCGC 1140  
 CAGATTAGCC GCGCGCGGCT CCGGCTCTCT ACTAGGCGC CCGCAATGCC CTCACCGGCT 1200  
 CTAAGCAGC CTTGCGCGCC TGGGCGGCG CTTGCGCGAT CAGGTGCTAG ATCCCGACAA 1260  
 CAGTCTCTT ACGCTCTAT CCGAAGCTT AAGCAAGCTT CTTCTCAAT CCGGCAAGT 1320  
 CAGCAAGCT CCGCTCTCT CCGAAGCTT CCGAAGCTT CCGAAGCTT CCGAAGCTT 1380  
 CCGCAAGCT CCGCTCTCT CCGAAGCTT CCGAAGCTT CCGAAGCTT CCGAAGCTT 1440  
 CAGCAAGCT CCGCTCTCT CCGAAGCTT CCGAAGCTT CCGAAGCTT CCGAAGCTT 1500  
 CCGAAGCT CCGCTCTCT CCGAAGCTT CCGAAGCTT CCGAAGCTT CCGAAGCTT 1560

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```

CTGCAGGGTG GCGTGGATGA GCGTCACCGC GGGGACGGCC GAGCTGACCG CCGCCAGGCT    60
CCGGGTGGCT GCGGUGGCGT ACCGACCGGC GTATGGCGTG ACCGTGCCCC CGCCGGTGAT    120
CGCCGAGAAC CGTGCTGAAC TGATGATTCT GATAGCGACC AACCTCTTGG GGC AAAACAC    180
CCCGCGGATC GCGGTCAACG AGGCCGAATA CGGCGAGATG TGGGCCCAAG ACGCCCGCGC    240
GATGTTTTGC TACGCGCGCG CGACGGCGAC GCGCAGCGCG ACGTTGCTG 13TTGAGGA    300
CGCCCGGGAG ATGACCAAGG CGGGTGGGCT GGTGAGGAG CGCGCGCGCG TCGAGGAGGG    360
CTCGACACCG CGCGCGGCGA ACCAGTTGAT GAAGAATGTG CGCGAGGCGG TGAAGAGATT    420
CGCCAGCGCG ACGGAGGCGA CGACCGCTTC TTGCAAGGTG GGTGGGCTGT GGAAGACGCT    480
CTCGCGCGAT CGGTGCGCGA TCAGCAACAT GGTGTGATG GCGAACAAGT ACATGTGAT    540
GACCAACTCG GGTGTGTGCA TGACCAACAC GTTGAGCTCG ATGTTGAAGG GCTTTGCTCC    600
CGCGCGCGCG CGCGAGGCGG TCAGAACCGT GCGCGAAAAA GGGGTGCGCG CGATGAGGTC    660
GCTGCGGAGC TCGGTGGGTT GTTGCGGCTT GGGCGGTGCG GTGCGCGCGA ACTTGCGTGG    720
CGCGCGCTCG GTAGGTATG GTACCGGCGA TCGCGGAAAA TATGCANAGT CTGCTGCGCG    780
GAAGGTGCTT CGGGGTAAAG GTTACCGCG GTTTGTGGA TCGGTGAAG TCGGTGAAG    840
GAAGAGTTTA C

```

1. INFORMATION FOR SEQ ID NO:34:

a. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 294 base pairs
- B. TYPE: nucleotide
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

a1. SEQUENCE DESCRIPTION: SEQ ID NO:34:

```

ATTGATCGG CGGAAAATT GAGGAGATT GGTGCGGT GATAAGGAA TGAATCGAG    60
TGAATTTAT TCGGTGAGG CGCGCGAGTA ATGCTGCGA CGAGAGAA GTTATGCTG    120
CGGAGCTG CGTATGTT TGAAGAGAA GAAAGGCTT GAAATTTT GAAAGAGAA    180

```

## (x) SEQUENCE CHARACTERISTICS.

- (A) LENGTH: 1227 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```

GATCCTGACC GAAGCGGCCG CCGCCAAGGC GAAGTCCCTG TTGGACCAGG AGGGACGGGA      60
CGATCTGGCG CTGCGGATCG CGGTTTCAAGC GGGGGGGTGC GCTGGATTGC GCTATAACCT      120
TTTCTTGGAC GACCGGACCG TGGATGGTGA CCAACCGCGC GAGTTCCGTG GTGTCAAGTT      180
GATCTGTGAC CGGATGAGCG TGGCGTATCT GGAAGGCGCG TGGATCGATT TCGTGGACAC      240
TATTGAGAAG GAAGGTTGAG GATCGACAAT CCAACCGCGC GCGGCTCTCT GCGCTCCCGG      300
GATTGCTTCA ACTGATAAAA CGGTAGTAGG ACCCGCGGCT GCGCAACAGG TACGAGTACA      360
CCAAAGAGTG ACCGCGCTGG AAAAGCAACT GAGCGATGCG TTGCACTTGA CCGGCTGGCG      420
GGCGCGCGCG GCGAGGTGTC ACCTGCATGG TGAACAGCAC CTGGGCTTGA TATTGCGACC      480
AGTACAGCAT TTTGTGGATC GAGGTCACTT CGAGCTGGGA GAACTGCTTG CGGAACGCGT      540
CGCTGCTTAC GTTGGCCAAG CGGTGATCGG AGCGCTTCTC GCGCAGCGCG TCGTGGATAC      600
CGCAGAGCGG ATTGCGAAGC ATGCTGTGCA GATCGCGCTT CTGAGCGCGG TTGAGGATAC      660
CTGGAATGCG GCTTTTGGCG GGTCCCTCGG AGAATCTGCG TGGCTGTCTG GCTGCTTTGG      720
TGGGAGCGCG GTATATGATG GCGCGCTTCA TACCGGACAG GAGCGGAGG GCTACGACAA      780
TGGCGATCAG CAGCGCGCTT TGGCTTGGCT TCGGCTAGGA GAGCTGCGCG GCGACGCGCG      840
GATATCGCGG GCGCGCGCAG TGGCGCTGCT TTGCGGCTCG GCGGCGGAG GCGGCTTGGG      900
GCGGAGAA GTGCTGCGG GATGCGAG GTGCTGCTT GTGCGATGAG GCGCTGCGG      960
GCGGAGAA GTGCTGCGG GATGCGAG GTGCTGCTT GTGCGATGAG GCGCTGCGG      1020
TTTCTTAGG GTGCTGCGG GAGCGAGCTT GTAGCGGAG AAGCGCGGCT GCGCTGAGCG      1080
GCGAGAGTGG GAAGTGAAGT GAGCTGCGTA GCGAGCTTAC GCGAGAGCT GCGCTGAGCT      1140
TTGAGTGGGA GCGGAGTGGT GCGGCGGCTT GTATATGATG AAGAGTACCG AAGCTTAGGA      1200
AGTAAAGAA GAGATTTCTT GAGATTTCTT

```

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```

GCGGTGTCCG CGGATCCGGC GGGTGGTTGA ACGGCAACGG CGGGGCCGGC GGGGCCGGCG      60
GGACCGGGCG TAACGGTGGT GCCGJCGGCA ACGCCTGGTT GTTCGGGGCC GCGCGTCCG      120
GCGGNGCCCG CACCAATGGT GGNGTCGGCG GTCGGGCGG ATTTGTCTAC GGCAACGGCG      180
1                                           191

```

2 INFORMATION FOR SEQ ID NO:37:

1. SEQUENCE CHARACTERISTICS:  
 A. LENGTH: 290 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```

GCGGTGTCCG CGGATCCGGC GGGTGGTTGA ACGGCAACGG CGGTGTCCG GCGCGGGGCG      60
GGACCGGGCG TAACGGTGGT GCCGJCGGCA ACGCCTGGTT GTTCGGGGCC GCGCGTCCG      120
GCGGNGCCCG CACCAATGGT GGNGTCGGCG GTCGGGCGG ATTTGTCTAC GGCAACGGCG      180
GGGTGACGGG CGGTTTCCGT GGGACCGGGG GTAAGGCTGG CGAGGGGCGG ATTGGGCGGG      240
GAGTGAAGAG CGGACCGGGT GTGGNGGTTG ACGGCGGTGA GGGCGGTGAC      290

```

3 INFORMATION FOR SEQ ID NO:38:

SEQUENCE CHARACTERISTICS  
 A. LENGTH: 14 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:38:

```

ATTGTAATGG CATGGNGGGT GTCAATGCAA GCAT      14

```

4 INFORMATION FOR SEQ ID NO:39:

1. SEQUENCE CHARACTERISTICS:  
 A. LENGTH: 155 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

GATGGCTGCT CGTCCCCCCC TTGCCCCCGA CCCCACCGGT CCCACCGTTA CCGAACAAGC 60  
 TGGCGTGGTC GCCAGCACCG CCGGCACCGC CGACCCCGGA GTCGAACAAT GGCACCGTCG 120  
 TATCCCCACC ATTGCCCCCG GNCCCCACCG CACCG 155

## (2) INFORMATION FOR SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 53 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATCCCCCTCA CCGGGCCCCG GGCACCGGC AGCCCCGNGG CCCCCGGGG TGG 60

## (2) INFORMATION FOR SEQ ID NO:41:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 131 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATTACCCG CGTTCAGAC CTTCCCCCG CCGCACCGG CACACCGGG CGTACCGGG 60  
 TATTTCCTT TAAGGCGCG AACCCACCG CCGTCGNGG CCCCCCGGG CCCCCGGGA 120  
 CCGCCGCAA TC 155

## (2) INFORMATION FOR SEQ ID NO:42:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 42 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATTTCCTT CGTACCGG CCGTACCGG TAAGCGCGG NAACCGCGG CCGTACCGG 60  
 TATTTCCTT ATCTTCGCG CCGTACCGG TTTTAAATG CCGTACCGG CCGTACCGG 120  
 TATTTCCTT TC 155

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO 43:

```

CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTCCCGGGG TTCCCCACC      60
CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCC CGATGCCGGC      120
ATCAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTTTCAGT TTAGCGACGA TAATGGGTAT      180
AGCACTAAGG AGGATGATCC GATATGACGC AGTCGCAGAC CGTGACGGTG GATCAGCAAG      240
AGATTTTGAA CAGGGCCAAC GAGGTGGAGG CCCCAGTGGC GGACCCACCG ACTGATGTCC      300
GCATCAGACC GTCCGAATC ACCGNGNTA AAAACCCCCC CCAACAGNTG GTNTGTCTCG      360
CCGACAACAT CCGCGAATAC GTCCCGCCCG GTGCCAAGA CGCGCAGCCT CTGGGACCT      420
CGGTGCGCAA CCGGGCCAAG GNGTATGGG AGGTTGATGA CGAGGCTGCG ACCGCGCTGG      480
ACAACGACCG CGAAGGAAC GTCCAGCCAC AATCGGCCCG GCGCGTCGGA GGGGACAGTT      540
CGGCGCAACT AACCGATAC CGGAGGCTCG CCACGGCCCG TGAACCCAAC TTCATGGATC      600
TGAAGAAGC CGCAAGGAAG CTCGAACCG CGCACCAAG CGCATCGCTC GCGCACTGNG      660
CGGATGGGCG GAACASTTC AGGCTGACCT TCGAAGCGCA CG                                702
  
```

(X2) INFORMATION FOR SEQ ID NO 44:

SEQUENCE CHARACTERISTICS:  
 A. LENGTH: 398 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

(X3) SEQUENCE DESCRIPTION: SEQ ID NO 44:

```

AAATGAGAG AGTGTCTCGG TCACTTATG TGAAGATG ATGTTCTCGT TCGGCTTGA      10
AGTTTGGG TCGGCTCGG AGGCTTATG GCGGATAT TGGGCTGCGA ATGCTGTGGT      20
TGGGCTGCGG GTGCTGACAT TATGCTGTTA TCGAAGAA TGGGCTGCGG GCGGCTGCGG      30
TGGGCTGCGG GTGCTGATCG AATGCTGATG GTGCTGCGG ATGAGGACA AGGCTGCGG      40
AACTTAAAT GTTCTGACAA TGAAGAAAT AGTCTGAT TGAAGATAT TGTGATG      50
  
```

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```

CGGCACGAGG ATCGAATCGC GTCGCCGGGA GCACAGCGTC GCACTGCACC AGTGGAGGAG      60
CCATGACCTA CTCGCCGGGT AACCCCGGAT ACCCGCAAGC GCAGCCCGCA GGCTCCTACG      120
GAGGCGTCAC ACCCTCGTTC GCCCAGCCCG ATGAGCGTGC GAGCAAGCTA CCGATGTACC      180
TGAACATCGC GGTGGCAGTG CTCGGTCTGG CTCCCTACTT CGCCAGCTTC GGCCCAATGT      240
TCACCTTCAG TACCGAACTC GGGGGGGGTG ATGGCCCACT GTCCGGTGAC ACTGGGCTGC      300
CGCTCCCGCT CGCTCTCTTC CTCGCCCTGC TTCCCGCGGT GGTTCCTGGT CCAAGGCCA      360
AGAGGCATGT GACGGTAGTT GCGGTGCTCG GGGTACTCGG CGTATTTCTG ATGGTCTCGG      420
CGAGCTTTAA CAAGCCCGAG CCTATTCTGA CGCGTTGGGC ATTGTGGCTT CTCTTGGCTT      480
TCATCTGTGT CCAGGCGGTT GCGGCAGTCC TGCGGCTGTT GGTGGAGACC GCGGCTATCA      540
CGCGCCCGGC CGCGCGGCCC AAGTTGAGC CCTATGGACA GTACGGGCGG TACCGGCACT      600
ACCGGCAGTA CGGGGTGCAG CGGGGTGGGT ACTACGGTCA GCAGGGTGCT CAGCAAGCCC      660
TGGCACTGCA GTCCGCCCGC CGGCAGCAGT CTCGCAGCC TCCCGGATAT GGGTCTAGT      720
ACCGCCGCTA TTCCTCCAGT CGGAGCCAAAT CGGCACTGG ATACACTGCT CAGGCCCGGG      780
CGCAGCCGCG GCGGCACTCG GGTTCGCAAC AATCGGACCA TGGCCCATCT ACGGCAGCTA      840
AGCTTTTTCG AGCTTCAGC GCACCAACAC CGCTCACTCG TGGAGCGGG TCGAGGCTG      900
ATCGGCTCG AGTCACTAT TCAAACCCCA CGGGGCGGCA CGAGTCTCG TCGCCCGGG      960
CGCGCGGCT GTAAACCGCG CTCGCCCGGT CTCTCGGCTT GGTTCGCGA AGAATGAAGA      1020
CGCTTATGAT GAGTTCGAG CATCTCTCTT TAAATT

```

(x1) INFORMATION FOR SEQ ID NO:46:

(A) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 327 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```

AGTGGCGCGG CGCGGCGGGG ACGGCGGCGG AGGCGGCGGT GGTGCGCTTC CAAGAAGCAG      180
CCAATAAGCA GAAGCAGGAA CTCGACGAGA TCTCGACGAA TATTCGTCAG GCGGCGCTCC      240
AATACTCGAG GGCCGACGAG GAGCAGCAGC AGGCGCTGTC CTCGCAAATG GGCTTCTGAC      300
CCGCTAATAC GAAAAGAAAC GGAGCAA                                           327

```

## (2) INFORMATION FOR SEQ ID NO:47:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 170 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## X1: SEQUENCE DESCRIPTION: SEQ ID NO:47:

```

GGTGGCGGAT GATGGCGCTG TCGAAGCTGA CGGATTCTGT ACCGCGCTCG TTCAGATCAA      60
CCAGCAACCT GTTGGCGCTG GCAAATGTGG CGNACCGCTG GATCTCGGTG ATCTTCTTCT      120
TCTTCATCAG GAAGTGACA CCGGCGACCG TGCGCTGGG TACCTTTTCG                      170

```

## (2) INFORMATION FOR SEQ ID NO:48:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## X1: SEQUENCE DESCRIPTION: SEQ ID NO:48:

```

GATCGCGCGG CACGCGCGGT GCGGCGCGCA CGACCGCTGG CGTGGCGGT AACGCGCGGG      60
CGCGCGGTGG CGCGCGAACC GTGCGCTTGG TCTTGGGCAA CGCGGTGGT GCGCGCGACG      120
CGCGCTT                                           127

```

## (2) INFORMATION FOR SEQ ID NO:49:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## X1: SEQUENCE DESCRIPTION: SEQ ID NO:49:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 149 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GATCAGGGCT GCGCGGCTCC GCGCAGAAGG GCGGTAACGG AGGAGCTGCC GGATTGTTTG 60  
 JCAACGGCGG GCGCGGNGGT GCGCGGCGCT CCAACCAAGC CGGTAACGGC GGNGCGCGGC 120  
 GAAACGGTGG TGCGGGTGGG CTGATCTGG 149

## (2) INFORMATION FOR SEQ ID NO:51:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGGCACGAGA TCACACCTAC CGACTGATCG AGATCGTCCG GACCTCGCCC GACGGTGTCTG 60  
 AGCGCGNAAT CGAGGGCGGT CTGGCGCGAG CTGCGCAGAC CATCGCGCCC CTGGACTGGT 120  
 TCGAAGTACA CTCGAATTCA GCGACGTCG TCGACCGAGC GGTGCGCGAC TTCCAGGTGA 180  
 CTATGAAAGT CGGCTTCGCG CTGGAGGATT CTGAAAGCTT CAGCGCGCGC CGATAACTGA 240  
 CTCGATCAT TAAGCGACTT TTGCAAAAGA CTCTAGCGCG CTCGAAAGCG GTTTCACCCG 300  
 AGGTTGGCTC CGCGGAGGCG CTCGCTCGAA AATCGTCTCG ACAATTCGTC GCGCG 355

## (2) INFORMATION FOR SEQ ID NO:52:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 399 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:52:

ATGATTAAG ATGAAAGTCA CAGCAGCAG CTGAAAGCA ATTGATA CTGAAAGTGA 60  
 CAGCAGCAG CTGAAAGTCA CAGCAGCAG CTGAAAGCA ATTGATA CTGAAAGTGA 120

```

GCCCCCGCCA ACACGCGGAA TCCCCAGCCG GCGGATCCCA ACGCAGCACC TCCGCGCGCC      300
GACCGGAACG CACCGCGCGC ACCTGTCATT GCGCCAAACG CACCCCAACC TGTCGGGATC      360
GACAACCCCG TTGGAGGATT CAGCTTCGCG CTGCTGCTG GCTGGGTGGA GTCTGACGCC      420
GCCCCACTTC ACTACGGTTC AGCACTCCTC AGCAAAACCA CCGGGGACCC GCCAATTCCC      480
GGACAGCCGC CGCCGGTGGC CAATGACACC CGTATCGTGC TCGGCCCGCT AGACCAAAAG      540
CTTTACGCCA GCGCCGAAGC CACCGACTCC AAGGCGCGCG CCGGGTTGGG CTCGGACATG      600
GGTGAGTTCT ATATGCCCTA CCGCGGCACC CGGATCAACC AGGAAACCGT CTCGCTCGAC      660
GCCAACGGGG TGTCTGGAAG CCGCTGCTAT TAGGAAGTCA AGTTCAGCGA TCCGAGTAAG      720
CCGAACCGCC AGATCTGGAC GGGCTAATC GGTCTGTCGG CCGCGAACCC ACCGGACGCC      780
GGGCCCCCTC AGCCTCGTT TGTGGTATGG CTGGGGACCG CCAACAACCC GGTGGACAAG      840
GGGCGCGCCA AGGCGCTGCC CGAATCGATC CCGCTTTTGG TCGCCCCGCC CCGGGCGCGC      900
GCACCGGCTC CTGCAGAGCC CGCTCCCGCG CCGGCGCGCG CCGGGGAAGT CGCTCTACC      960
CGGACGACAC CGACACCGCA CCGGACCTTA CCGGGCTGA      999

```

# 2. INFORMATION FOR SEQ ID NO.13:

1. SEQUENCE CHARACTERISTICS:
  - A. LENGTH: 332 amino acids
  - B. TYPE: amino acid
  - C. STRANDEDNESS: single
  - D. TOPOLOGY: linear

## 3. SEQUENCE DESCRIPTION: SEQ ID NO.13

```

Met His His His His His His Met His Gln Val Asp Pro Asn Leu Thr
1           5           10           15
Asp Ala Val Ile Arg Val His Val Val Ile Val Val Met Ala Ser
21           25           30           35
Asp Thr Val Thr Thr Val Val Thr Thr Val Thr Val Asn Ala Asp Pro
41           45           50           55
His Thr Ala Thr Pro Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
61           65           70           75
Asp Ala Ala Thr Thr Thr Thr Thr Thr Thr Val Ala Thr Thr Thr Thr
81           85           90           95

```

100	105	110
Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser 115 120 125		
Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp 130 135 140		
Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro 145 150 155 160		
Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg 165 170 175		
Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala 180 185 190		
Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro 195 200 205		
Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val 210 215 220		
Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys 225 230 235 240		
Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn 245 250 255		
Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly 260 265 270		
Thr Ala Asn Asn Pro Val Asp Lys Ile Ala Ala Lys Ala Leu Ala Glu 275 280 285		
Ser Ile Arg Pro Leu Val Ala Pro Trp Pro Ala Pro Ala Pro Ala Pro 290 295 300		
Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Ala Ala Ala Ala Ala 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400		

# INFORMATION FOR SEQ ID NO:84

## 1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 401 amino acids
- B. TYPE: protein
- C. STRANDEDNESS: single

2) INFORMATION FOR SEQ ID NO:55:

X1 SEQUENCE DESCRIPTION: SEQ ID NO:55:

2 INFORMATION FOR SEQ ID NO: 55

24 SEQUENCE DESCRIPTION. SEQ ID NO:55

THE JAY ARD

INFORMATION FOR SEQ ID NO: 57

THESE RESULTS ARE IN ACCORD WITH THE

DECLASSIFICATION AUTHORITY: 25X, 25Y, 25Z

1. **THEORY** 2. **EXPERIMENT** 3. **CONCLUSION**

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val  
1 5 10

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Ala Glu Gln Ser Ile Ser Thr Xaa Gln Xaa Ile Val Pro  
1 5 10

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Ala Ala Ala Ala Pro Pro  
1 5 10 15

Ala

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Pro Lys Thr Tyr Xaa Ile Gly Leu Lys Gly Thr Asp Thr Gly  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Gln Thr Ser  
 1 5 10 15  
 Leu Leu Asn Asn Leu Ala Asp Pro Asp Val Ser Phe Ala Asp  
 20 25 30

## (2) INFORMATION FOR SEQ ID NO:63:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## x1 SEQUENCE DESCRIPTION: SEQ ID NO:63:

Gly Cys Gly Asp Arg Ser Gly Gly Asn Leu Asp Gln Ile Arg Leu Arg  
 1 5 10 15  
 Arg Asp Arg Ser Gly Gly Asn Leu  
 20

## 2 INFORMATION FOR SEQ ID NO:64:

## 1 SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 187 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## x1 SEQUENCE DESCRIPTION: SEQ ID NO:64:

Thr Gly Ser Leu Asn Gln Thr His Asn Arg Arg Ala Asn Gln Arg Cys  
 1 5 10 15  
 Asn Thr Thr Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala  
 20 25 30  
 Asn Val Lys Lys Lys Ala Ala Ala Lys Gln Val Thr Ser Ile Met Ala  
 35 40 45  
 Gly Gly Pro Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro  
 50 55 60  
 Leu Pro Leu Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln  
 65 70 75 80  
 Leu Thr Ser Ile Leu Asp Ser Leu Ala Asp Ser Thr Val Ser Thr Ala

```

Ile Ala Asp His Lys Leu Lys Lys Ala Ala Glu His Gly Asp Leu Pro
  115                      120                      125

Leu Ser Phe Ser Val Thr Asn Ile Gln Pro Ala Ala Ala Gly Ser Ala
  130                      135                      140

Thr Ala Asp Val Ser Val Ser Gly Pro Lys Leu Ser Ser Pro Val Thr
  145                      150                      155                      160

Gln Asn Val Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala
  165                      170                      175

Ser Ala Met Glu Leu Leu Gln Ala Ala Gly Xaa
  180                      185

```

2. INFORMATION FOR SEQ ID NO:65:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 148 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:65:

```

Asp Glu Val Thr Val Glu Thr Thr Ser Val Phe Arg Ala Asp Phe Leu
  1      5      10      15

Ser Gln Leu Asp Ala Pro Ala Gln Ala Gly Thr Glu Ser Ala Val Ser
  20      25      30

Gly Val Gln Gly Leu Pro Pro Gly Ser Ala Leu Leu Val Val Lys Arg
  35      40      45

Gly Pro Asn Ala Gly Ser Arg Phe Leu Leu Asp Gln Ala Ile Thr Ser
  50      55      60

Ala Gly Arg His Pro Asp Ser Asp Ile Phe Leu Asp Asp Val Thr Val
  65      70      75      80

Ser Arg Arg His Val Thr Phe Arg Leu Gln Asn Asn Thr Phe Asn Val
  85      90      95

Val Asp Val Gly Ser Thr Asn Gly Thr Thr Val Asn Arg Gln Pro Val
  100     105     110     115

Asp Ser Ala Val Leu Ala Asn Gly Asp Gln Val Gln Ile His Lys Leu
  120     125     130     135

Arg Leu Val Phe Leu Thr Gly Pro Asn Gly Thr Val Asp Asp His Ser
  140     145     150     155

```

## (2) INFORMATION FOR SEQ ID NO:66:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 230 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```

Thr Ser Asn Arg Pro Ala Arg Arg Gly Arg Arg Ala Pro Arg Asp Thr
1           5           10           15

Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln
20           25           30

Gln Arg Asp Ala Leu Cys Leu Ser Ser Thr Gln Ile Ser Arg Gln Ser
35           40           45

Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn
50           55           60

Phe Asp Val Arg Ile Lys Ile Phe Met Leu Val Thr Ala Val Val Leu
65           70           75           80

Leu Cys Cys Ser Gly Val Ala Thr Ala Ala Pro Lys Thr Tyr Cys Glu
85           90           95

Glu Leu Lys Gly Thr Asp Thr Gly Gln Ala Cys Gln Ile Gln Met Ser
100          105          110

Asp Pro Ala Tyr Asn Ile Asn Ile Ser Leu Pro Ser Tyr Tyr Pro Asp
115          120          125

Ile Lys Ser Leu Glu Asn Tyr Ile Ala Gln Thr Arg Asp Lys Phe Leu
130          135          140

Ser Ala Ala Thr Ser Ser Thr Pro Arg Glu Ala Pro Tyr Glu Leu Asn
145          150          155          160

Leu Thr Ser Ala Thr Tyr Gln Ser Ile Ile Pro Ser Arg Gly Thr Ile
165          170          175

Ile Glu Val Leu Asn Val Cys His Asn Ala Gly Gly Thr His Pro Thr
180          185          190

Thr Thr Thr Lys Ala Phe Asp Trp Asp Gln Ala Tyr Arg Lys Pro Ile
195          200          205

Thr Thr Asp Thr Ser Thr Gln Ala Asn Thr Asp Pro Leu Pro Val Val

```

## 2) INFORMATION FOR SEQ ID NO:67:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```

Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
 1             5             10             15

Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
 20             25             30

Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
 35             40             45

Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
 50             55             60

Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
 65             70             75             80

Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
 85             90             95

Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp
100            105            110

Ile Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Gln
115            120            125

Gly Pro Pro Ala
130

```

## 2) INFORMATION FOR SEQ ID NO:68:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:68:

```

Val Pro Leu Arg Ser Pro Ser Met Ser Pro Ser Lys Thr Leu Ala Ala
 1             5             10

```

Met Ala Arg Val Arg Arg Arg Ala Ile Trp Arg Gly Pro Ala Thr Xaa  
 50 55 60

Ser Ala Gly Met Ala Arg Val Arg Arg Trp Xaa Val Met Pro Xaa Val  
 65 70 75 80

Ile Gln Ser Thr Xaa Ile Arg Xaa Xaa Gly Pro Phe Asp Asn Arg Gly  
 85 90 95

Ser Glu Arg Lys  
 100

# 2. INFORMATION FOR SEQ ID NO:69:

## (A) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 163 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (X) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Glu Arg Val Arg Thr  
 1 5 10 15

Leu Thr Leu Asn Arg Pro Gln Ser Arg Asn Ala Leu Ser Ala Ala Leu  
 20 25 30

Arg Asp Arg Phe Phe Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp  
 35 40 45

Ile Asp Val Val Ile Leu Thr Gly Ala Asp Pro Val Thr Phe Ala Thr  
 50 55 60

Leu Asp Leu Lys Val Ala Gly Arg Ala Asp Arg Ala Val Gly His Leu  
 65 70 75 80

Thr Ala Val Leu His Asp Val Ala His Asp Arg Arg Asp Lys Arg  
 85 90 95

Arg Arg Leu Ala Arg Arg Ser Ala Thr Val Glu Leu Arg Ala Arg  
 100 105 110

Asp Arg Leu Arg Ala Arg Pro Leu Arg Arg His Pro Arg Pro Gly Gly  
 115 120 125

Ala Ala Ala His Leu Gly Thr Thr Gly Val Leu Ala Ala Lys Gly Arg  
 130 135 140

## 2 INFORMATION FOR SEQ ID NO:70:

## 1' SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 344 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix1) SEQUENCE DESCRIPTION: SEQ ID NO:70:

```

Met Lys Phe Val Asn His Ile Glu Pro Val Ala Pro Arg Arg Ala Gly
 1             5             10             15

Gly Ala Val Ala Glu Val Tyr Ala Glu Ala Arg Arg Glu Phe Gly Arg
 20             25             30

Leu Pro Glu Pro Leu Ala Met Leu Ser Pro Asp Glu Gly Leu Leu Thr
 35             40             45

Ala Gly Trp Ala Thr Leu Arg Glu Thr Leu Leu Val Gly Gln Val Pro
 50             55             60

Arg Gly Arg Lys Glu Ala Val Ala Ala Ala Val Ala Ala Ser Leu Arg
 65             70             75             80

Tys Pro Trp Cys Val Asp Ala His Thr Thr Met Leu Tyr Ala Ala Gly
 85             90             95

Gln Thr Asp Thr Ala Ala Ala Ile Leu Ala Gly Thr Ala Pro Ala Ala
100            105            110

His Asp Pro Asn Ala Pro Tyr Val Ala Thr Ala Ala Gly Thr Gly Thr
115            120            125

Pro Ala Gly Pro Pro Ala Pro Phe Gly Pro Asp Val Ala Ala Glu Tyr
130            135            140

Ser Glu Thr Ala Val Glu Phe His Phe Ile Ala Arg Leu Val Ser Val
145            150            155            160

Leu Leu Asp Glu Thr Ser Ser Val Glu Ser Ser Arg Val Ala Glu Leu
165            170            175            180

Met Arg Arg Ala Gly Gly Leu Val Phe Ala Arg Lys Val Arg Ala Glu
185            190            195

His Arg Pro His Arg Ser Thr Arg Arg Leu His Pro Arg Thr Leu Pro
200            205            210

```

Thr Arg Gln Val Val Arg Arg Val Val Gly Ser Trp His Gly Glu Pro  
 245 250 255  
 Met Pro Met Ser Ser Arg Trp Thr Asn Glu His Thr Ala Glu Leu Pro  
 260 265 270  
 Ala Asp Leu His Ala Pro Thr Arg Leu Ala Leu Leu Thr Gly Leu Ala  
 275 280 285  
 Pro His Gln Val Thr Asp Asp Asp Val Ala Ala Ala Arg Ser Leu Leu  
 290 295 300  
 Asp Thr Asp Ala Ala Leu Val Gly Ala Leu Ala Trp Ala Ala Phe Thr  
 305 310 315 320  
 Ala Ala Arg Arg Ile Gly Thr Trp Ile Gly Ala Ala Ala Glu Gly Gln  
 325 330 335  
 Val Ser Arg Gln Asn Pro Thr Gly  
 340

## 2 INFORMATION FOR SEQ ID NO:71:

### 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 485 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### 2. SEQUENCE DESCRIPTION: SEQ ID NO:71

Asp Asp Pro Asp Met Pro Gly Thr Val Ala Lys Ala Val Ala Asp Ala  
 1 5 10  
 Leu Gly Arg Gly Ile Ala Pro Val Gln Asp Ile Gln Asp Lys Val Glu  
 15 20 25 30  
 Ala Arg Leu Gly Gln Ala Gly Ile Ser Asp Val Ala Arg Val Tyr Ile  
 35 40 45 50  
 Leu Thr Arg Gln Arg Arg Ala Ile Val Arg Thr Ala Val Ala Leu Leu  
 55 60 65  
 Gln Val Arg Asp Gln Leu Lys Leu Ser Leu Ala Ala Val Thr Val Leu  
 70 75 80  
 Arg Gln Arg Tyr Leu Leu His Asp Gln Gln Gly Arg Pro Ala Val Ser  
 85 90 95

Thr Leu Leu Arg Asn Leu Glu Phe Leu Pro Asn Ser Pro Thr Leu Met  
 130 135 140

Asn Ser Gly Thr Asp Leu Gly Leu Leu Ala Gly Cys Phe Val Leu Pro  
 145 150 155 160

Ile Glu Asp Ser Leu Gln Ser Ile Phe Ala Thr Leu Gly Gln Ala Ala  
 165 170 175

Glu Leu Gln Arg Ala Gly Gly Gly Thr Gly Tyr Ala Phe Ser His Leu  
 180 185 190

Arg Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly  
 195 200 205

Pro Val Ser Phe Leu Arg Leu Tyr Asp Ser Ala Ala Gly Val Val Ser  
 210 215 220

Met Gly Gly Arg Arg Arg Gly Ala Cys Met Ala Val Leu Asp Val Ser  
 225 230 235 240

His Pro Asp Ile Cys Asp Phe Val Thr Ala Lys Ala Glu Ser Pro Ser  
 245 250 255

Glu Leu Pro His Phe Asn Leu Ser Val Gly Val Thr Asp Ala Phe Leu  
 260 265 270

Arg Ala Val Glu Arg Asn Gly Leu His Arg Leu Val Asn Pro Arg Thr  
 275 280 285

Gly Lys Ile Val Ala Arg Met Pro Ala Ala Glu Leu Phe Asp Ala Ile  
 290 295 300

Lys Lys Ala Ala His Ala Gly Gly Asp Pro Gly Leu Val Phe Leu Asp  
 305 310 315 320

Thr Ile Asn Arg Ala Asn Pro Val Pro Gly Arg Gly Arg Ile Glu Ala  
 325 330 335

Thr Asn Pro Lys Gly Glu Val Pro Leu Leu Pro Tyr Ile Ser Val Asn  
 340 345 350

Met Gly Ser Ile Asn Leu Ala Arg Met Leu Ala Asp Gly Arg Val Asp  
 355 360 365

Asp Asp Arg Leu Glu Glu Val Ala Gly Val Ala Val Arg Phe Leu Asp  
 370 375 380

Asn Val Ile Asp Val Ser Arg Tyr Pro Phe Ile Glu Leu Val Ala  
 385 390 395

Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg  
 420 425 430  
 Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala  
 435 440 445  
 Ser Arg Arg Leu Ala Glu Glu Arg Gly Ala Phe Pro Ala Phe Thr Asp  
 450 455 460  
 Ser Arg Phe Ala Arg Ser Gly Pro Arg Arg Asn Ala Gln Val Thr Ser  
 465 470 475 480  
 Val Ala Pro Thr Gly  
 485

(xii) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 267 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xiii) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Gly Val Ile Val Leu Asp Leu Glu Pro Arg Gly Pro Leu Pro Thr Glu  
 1 5 10 15  
 Ile Tyr Trp Arg Arg Arg Gly Leu Ala Leu Gly Ile Ala Val Val Val  
 20 25 30  
 Val Gly Ile Ala Val Ala Ile Val Ile Ala Phe Val Asp Ser Ser Ala  
 35 40 45  
 Gly Ala Lys Pro Val Ser Ala Asp Lys Pro Ala Ser Ala Gln Ser His  
 50 55 60  
 Pro Gly Ser Pro Ala Pro Gln Ala Pro Gln Pro Ala Gly Gln Thr Glu  
 65 70 75 80  
 Ile Asn Ala Ala Ala Ala Pro Arg Gln Ile Gln Asn Pro Gln Thr Pro  
 85 90 95  
 Thr Pro Thr Ala Ala Val Val Ala Pro Arg Val Leu Lys Gln Gly Asp  
 100 105 110  
 Asp Lys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro  
 115 120 125  
 Gln Tyr Tyr Val Val Asn Gln Pro Lys Phe Thr Met Val Val Thr Asn  
 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000

Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp  
 165 170 175

Cys Ala Pro Ser Asn Glu Thr Leu Val Lys Thr Phe Ser Pro Gly Glu  
 180 185 190

Gln Val Thr Thr Ala Val Thr Trp Thr Gly Met Gly Ser Ala Pro Arg  
 195 200 205

Cys Pro Leu Pro Arg Pro Ala Ile Gly Pro Gly Thr Tyr Asn Leu Val  
 210 215 220

Val Gln Leu Gly Asn Leu Arg Ser Leu Pro Val Pro Phe Ile Leu Asn  
 225 230 235 240

Gln Pro Pro Pro Pro Gly Pro Val Pro Ala Pro Gly Pro Ala Gln  
 245 250 255

Ala Pro Pro Pro Glu Ser Pro Ala Gln Gly Gly  
 260 265

# 1. INFORMATION FOR SEQ ID NO:73:

## (i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 97 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly Val Gln Val  
 1 5 10 15

Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Gln Val Val Ala  
 20 25 30

Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val Val Val Thr  
 35 40 45

Asp Val Asp Asp Asp Thr Ile Asn Ser Ala Ser Ala Leu Val Ala Ala  
 50 55 60

Val Asp Ser Lys Ala Pro Gly Ala Thr Val Ala Ser Thr Pro His Asp  
 65 70 75 80

Pro Ser Gly Gly Ser Asp Thr Val Gln Val Thr Leu Gly Lys Ala Gln  
 85 90 95

Gln

- (A) LENGTH: 364 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Gly Ala Ala Val Ser Leu Leu Ala Ala Gly Thr Leu Val Leu Thr Ala  
 1 5 10 15  
 Cys Gly Gly Gly Thr Asn Ser Ser Ser Ser Gly Ala Gly Gly Thr Ser  
 20 25 30  
 Gly Ser Val His Cys Gly Gly Lys Lys Glu Leu His Ser Ser Gly Ser  
 35 40 45  
 Thr Ala Gln Glu Asn Ala Met Glu Gln Phe Val Tyr Ala Tyr Val Arg  
 50 55 60  
 Ser Cys Pro Gly Tyr Thr Leu Asp Tyr Asn Ala Asn Gly Ser Gly Ala  
 65 70 75 80  
 Gly Val Thr Gln Phe Leu Asn Asn Glu Thr Asp Phe Ala Gly Ser Asp  
 85 90 95  
 Val Pro Leu Asn Pro Ser Thr Gly Gln Pro Asp Arg Ser Ala Glu Arg  
 100 105 110  
 Cys Gly Ser Pro Ala Trp Asp Leu Pro Thr Val Phe Gly Pro Ile Ala  
 115 120 125  
 Ile Thr Tyr Asn Ile Lys Gly Val Ser Thr Leu Asn Leu Asp Gly Pro  
 130 135 140  
 Thr Thr Ala Lys Ile Phe Asn Gly Thr Ile Thr Val Trp Asn Asp Pro  
 145 150 155 160  
 Gln Ile Gln Ala Leu Asn Ser Gly Thr Asp Leu Pro Pro Thr Pro Ile  
 165 170 175  
 Ser Val Ile Phe Asn Ser Asp Lys Ser Thr Thr Ser Asp Asn Pro Gln  
 180 185 190  
 Lys Tyr Leu Asp Gly Ala Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser  
 195 200 205  
 Glu Thr Phe Ser Gly Gly Val Gly Val Gly Ala Ser Gly Asn Asn Gly  
 210 215 220  
 Thr Ser Ala Leu Leu Gln Thr Thr Ala Glu Ser Ile Thr Thr Asn Gln

Ser Ala Gly Pro Asp Pro Val Ala Ile Thr Thr Glu Ser Val Gly Lys  
 260 265 270  
 Thr Ile Ala Gly Ala Lys Ile Met Gly Gln Gly Asn Asp Leu Val Leu  
 275 280 285  
 Asp Thr Ser Ser Phe Tyr Arg Pro Thr Gln Pro Gly Ser Tyr Pro Ile  
 290 295 300  
 Val Leu Ala Thr Tyr Glu Ile Val Cys Ser Lys Tyr Pro Asp Ala Thr  
 305 310 315 320  
 Thr Gly Thr Ala Val Arg Ala Phe Met Gln Ala Ala Ile Gly Pro Gly  
 325 330 335  
 Gln Glu Gly Leu Asp Gln Tyr Gly Ser Ile Pro Leu Pro Lys Ser Phe  
 340 345 350  
 Gln Ala Lys Leu Ala Ala Ala Val Asn Ala Ile Ser  
 355 360

# (1) INFORMATION FOR SEQ ID NO:75:

## (A) SEQUENCE CHARACTERISTICS:

- (1) LENGTH: 369 amino acids
- (2) TYPE: amino acid
- (3) STRANDEDNESS: single
- (4) TOPOLOGY: linear

## (X) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Gln Ala Ala Ala Gly Arg Ala Val Arg Arg Thr Gly His Ala Gln Asp  
 1 5 10 15  
 Gln Thr His Gln Asp Arg Leu His His Gly Cys Arg Arg Ala Ala Val  
 20 25 30  
 Val Val Arg Gln Asp Arg Ala Ser Val Ser Ala Thr Ser Ala Arg Pro  
 35 40 45  
 Pro Arg Arg His Pro Ala His Ile His Arg Arg Arg Val Ala Ile Ser  
 50 55 60  
 Gln Gly Arg Arg Arg Pro His Pro His His Val Gln Pro Asp Asp Arg  
 65 70 75 80  
 Arg Asp Arg Pro Ala Leu Leu Asp Arg Thr Gln Pro Ala Gln His Pro  
 85 90 95  
 Asn Pro His Arg Ala Glu Pro Ala Asp Pro Gly Arg Val Arg Gln Arg

Ala Asp His Gly Ala Pro Val Arg Gly Arg Gly Pro His Arg Gly Val  
 130 135 140

Gln His Arg Gly Gly Pro Val Phe Val Arg Arg Val Pro Gly Val Arg  
 145 150 155 160

Cys Ala His Arg Arg Gly His Arg Arg Val Ala Ala Pro Gly Gln Gly  
 165 170 175

Asp Val Leu Arg Ala Gly Leu Arg Val Glu Arg Leu Arg Pro Val Ala  
 180 185 190

Ala Val Glu Asn Leu His Arg Gly Ser Gln Arg Ala Asp Gly Arg Val  
 195 200 205

Phe Arg Pro Ile Arg Arg Gly Ala Arg Leu Pro Ala Arg Arg Ser Arg  
 210 215 220

Ala Gly Pro Gln Gly Arg Leu His Leu Asp Gly Ala Gly Pro Ser Pro  
 225 230 235 240

Leu Pro Ala Arg Ala Gly Gln Gln Gln Pro Ser Ser Ala Gly Gly Arg  
 245 250 255

Arg Ala Gly Gly Ala Glu Arg Ala Asp Pro Gly Gln Arg Gly Arg His  
 260 265 270

His Gln Gly Gly His Asp Pro Gly Arg Gln Gly Ala Gln Arg Gly Thr  
 275 280 285

Ala Gly Val Ala His His Ala Ala Gly Pro Arg Arg Ala Ala Val Arg  
 290 295 300

Leu Arg Pro Arg Arg  
 305

# 2. INFORMATION FOR SEQ ID NO:16

## 2.1 SEQUENCE CHARACTERISTICS

- A LENGTH: 30 amino acids
- B TYPE: amino acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

## 2.2 SEQUENCE DESCRIPTION: SEQ ID NO: 16

Leu Ala Val Thr Tyr Leu Asn Gly Pro Thr Gln Arg His Arg His Gly  
 1 10 20

Arg Gln Arg Val Ser Ala Ser Gln Thr Arg Ser Ser Asn Arg Thr Tyr  
 21 30

Ser Pro Leu Glu Arg Arg Phe Thr Cys Cys Ser Pro Ala Val Gly Cys  
 50 55 60

Arg Phe Arg Ser Phe Pro Val Arg Arg Leu Ala Leu Gly Ala Arg Thr  
 65 70 75 80

Ser Arg Thr Leu Gly Val Arg Arg Thr Leu Ser Gln Trp Asn Leu Ser  
 85 90 95

Pro Arg Ala Gln Pro Ser Cys Ala Val Thr Val Glu Ser His Thr His  
 100 105 110

Ala Ser Pro Arg Met Ala Lys Leu Ala Arg Val Val Gly Leu Val Gln  
 115 120 125

Glu Glu Gln Pro Ser Asp Met Thr Asn His Pro Arg Tyr Ser Pro Pro  
 130 135 140

Pro Gln Gln Pro Gly Thr Pro Gly Tyr Ala Gln Gly Gln Gln Gln Thr  
 145 150 155 160

Tyr Ser Gln Gln Phe Asp Trp Arg Tyr Pro Pro Ser Pro Pro Pro Gln  
 165 170 175

Pro Thr Gln Tyr Arg Gln Pro Tyr Glu Ala Leu Gly Gly Thr Arg Pro  
 180 185 190

Gly Leu Ile Pro Gly Val Ile Pro Thr Met Thr Pro Pro Pro Gly Met  
 195 200 205

Val Arg Gln Arg Pro Arg Ala Gly Met Leu Ala Ile Gly Ala Val Thr  
 210 215 220

Ile Ala Val Val Ser Ala Gly Ile Thr Gly Ala Ala Ala Ser Leu Val  
 225 230 235 240

Gly Phe Asn Arg Ala Pro Ala Gly Pro Ser Gly Gly Pro Val Ala Ala  
 245 250 255

Val Ala Ala Thr Val Thr Val Ala Ala Asn Met Ser Thr Gly Ser Val  
 260 265

Glu Gln Val Ala Ala Leu Val Val Thr Ser Val Thr Met Leu Thr Thr  
 270 275 280

Asp Leu Gly Arg Gln Ser Glu Gln Gly Ser Gly Ile Ile Leu Ser Ala  
 285 290 295 300

Val Ser Ser Thr Thr Thr Asn Asn His Val Val Ala Ala Val Val Lys  
 305 310 315 320

340	345	350
Ile Ala Val Val Arg Val Gln Gly Val Ser Gly Leu Thr Pro Ile Ser		
355	360	365
Leu Gly Ser Ser Ser Asp Leu Arg Val Gly Gln Pro Val Leu Ala Ile		
370	375	380
Gly Ser Pro Leu Gly Leu Glu Gly Thr Val Thr Thr Gly Ile Val Ser		
385	390	395
Ala Leu Asn Arg Pro Val Ser Thr Thr Gly Glu Ala Gly Asn Gln Asn		
405	410	415
Thr Val Leu Asp Ala Ile Gln Thr Asp Ala Ala Ile Asn Pro Gly Asn		
420	425	430
Ser Gly Gly Ala Leu Val Asn Met Asn Ala Gln Leu Val Gly Val Asn		
435	440	445
Ser Ala Ile Ala Thr Leu Gly Ala Asp Ser Ala Asp Ala Gln Ser Gly		
450	455	460
Ser Ile Gly Leu Gly Phe Ala Ile Pro Val Asp Gln Ala Lys Arg Ile		
465	470	475
Ala Asp Gln Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly		
485	490	495
Val Gln Val Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Gln		
500	505	510
Ala Val Ala Gln Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val		
515	520	525
Val Val Thr Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu		
530	535	540
Val Ala Ala Val Arg Ser Ser Ala Ser Ile Ala Thr Val Ala Leu Thr		
545	550	555
Arg Val Arg Val Ser Gly Leu Ser Arg Thr Val Gln Val Thr Leu Val		
560	565	570
Lys Ala Gln Gln		
580		

1. INFORMATION FOR SEQ ID NO 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 580

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

```

Met Asn Asp Gly Lys Arg Ala Val Thr Ser Ala Val Leu Val Val Leu
 1             5             10             15

Gly Ala Cys Leu Ala Leu Trp Leu Ser Gly Cys Ser Ser Pro Lys Pro
 20             25             30

Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr Ala Ser Asp Pro
 35             40             45

Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala Thr Lys Gly Leu
 50             55             60

Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys Val Asp Ser Leu
 65             70             75             80

Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala Asn Pro Leu Ala
 85             90             95

Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly Val Pro Phe Arg
100             105             110

Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp Asp Trp Ser Asn
115             120             125

Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala
130             135             140

Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln
145             150             155             160

Gly Thr His Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Glu Thr
165             170             175

Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly Ala Lys Ser Ala
180             185             190

Arg Pro Ala Thr Val Thr Ile Val Val Val Val Ser His His Leu Val
195             200             205             210

Val Val Ser Thr Asp Val Val Val Val Val Val Val Val Val Val Val
215             220             225             230

Lys Thr Asn Glu Pro Val Asn Val Asp
235             240

```

INFORMATION FOR SEQ ID NO:77:

SEQUENCE CHARACTERISTICS

ix1. SEQUENCE DESCRIPTION: SEQ ID NO:78:

Val Ile Asp Ile Ile Gly Thr Ser Pro Thr Ser Trp Glu Gln Ala Ala  
 1 5 10 15  
 Ala Glu Ala Val Gln Arg Ala Arg Asp Ser Val Asp Asp Ile Arg Val  
 20 25 30  
 Ala Arg Val Ile Glu Gln Asp Met Ala Val Asp Ser Ala Gly Lys Ile  
 35 40 45  
 Thr Tyr Arg Ile Lys Leu Glu Val Ser Phe Lys Met Arg Pro Ala Gln  
 50 55 60  
 Pro Arg  
 65

2. INFORMATION FOR SEQ ID NO:78:

1. SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 69 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

xi. SEQUENCE DESCRIPTION: SEQ ID NO:79:

Val Pro Pro Ala Pro Pro Leu Pro Pro Leu Pro Pro Ser Pro Ile Ser  
 1 5 10 15  
 Lys Ala Ser Pro Pro Ser Pro Pro Leu Pro Pro Ala Pro Pro Val Ala  
 20 25 30  
 Pro Ile Pro Pro Met Pro Pro Leu Asp Pro Trp Pro Pro Ala Pro Pro  
 35 40 45  
 Leu Pro Tyr Ser Thr Pro Pro Gly Ala Pro Leu Pro Pro Ser Pro Pro  
 50 55 60  
 Pro Trp Trp Lys Lys  
 65

3. INFORMATION FOR SEQ ID NO:79:

1. SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 65 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

Val Leu Ala Ala Val Gly Leu Gly Leu Ala Thr Ala Pro Ala Gln Ala  
 20 25 30  
 Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu  
 35 40 45  
 Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Ala Pro Gln Val Val  
 50 55 60  
 Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr  
 65 70 75 80  
 Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val  
 85 90 95  
 Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln  
 100 105 110  
 Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala  
 115 120 125  
 Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly  
 130 135 140  
 Gly Gly Val Ala Val Gly Gln Pro Val Val Ala Met Gly Asn Ser Gly  
 145 150 155 160  
 Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu  
 165 170 175  
 Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Gln Gly Thr  
 180 185 190  
 Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser  
 195 200 205  
 Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr  
 210 215 220  
 Ala Ala Thr Asn Asn Phe His Leu Pro His His Gly Ser Gly Thr Ala  
 225 230 235 240  
 Ile Pro Gly Gly Gln Ala Val Ala His Ala Gly Gln Ile Arg Ser Gly  
 245 250 255  
 Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu  
 260 265 270  
 Gly Val Val Asn Asn Asn Gly Asn His Ala Asn Val Gln Arg Val Val  
 275 280 285

305                      310                      315                      320  
 Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln  
                          325                      330                      335  
 Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly  
                          340                      345                      350  
 Pro Pro Ala  
                          355

## (2) INFORMATION FOR SEQ ID NO:81:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 305 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Ser Pro Lys Pro Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr  
 1                      5                      10                      15  
 Ala Ser Asp Pro Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala  
                          20                      25                      30  
 Thr Lys Gly Leu Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys  
                          35                      40                      45  
 Val Asp Ser Leu Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala  
                          50                      55                      60  
 Asn Pro Leu Ala Ala Lys Gly Val Lys Thr Tyr Asn Asp Gln Gln Ile  
                          65                      70                      75                      80  
 Val Pro Phe Arg Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp  
                          85                      90                      95  
 Val Thr Ser Asn Leu Thr Ser Gln Ser Gln Leu Ser Thr Ser Val Val  
                          100                      105                      110  
 Asn Asn Pro Val Ala Glu Thr Thr Asn Leu Leu Ser Gly Val Thr Asn  
                          115                      120                      125  
 Leu Gln Ala Gln Gly Thr Gln Val Ile Asp Gly Ile Ser Thr Thr Lys  
                          130                      135                      140  
 Ile Thr Ile Thr Ile Pro Ala Ser Ser Val Lys Met Leu Asn Gln Gln  
                          145                      150                      155                      160

180

185

190

Leu Thr Gln Ser Lys Trp Asn Glu Pro Val Asn Val Asp  
 195 200 205

## (2) INFORMATION FOR SEQ ID NO:82:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 286 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val  
 1 5 10 15  
 Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln  
 20 25 30  
 His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val  
 35 40 45  
 Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Lys Glu  
 50 55 60  
 Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe  
 65 70 75 80  
 Tyr Ile Thr Val Tyr Asn Gln Pro Tyr Val Gln Pro Pro Glu Pro Glu  
 85 90 95  
 Asn Phe Asp Pro Gln Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala  
 100 105 110  
 Ala Thr Gln Gln Arg Thr Asn Lys Lys Gln Ile Leu Ala Ser Gly Val  
 115 120 125  
 Val Met Pro Ala Ala Leu Arg Ala Val Gln Met Leu Ala Val Tyr  
 130 135 140  
 Asp Val Ala Val Asp Ala Thr Ser Val Thr Ser Thr Val Glu Val Asp  
 145 150 155 160  
 Arg Asn Gly Thr Val Ile Gln Thr Gln Lys Ser Arg His Pro Asn Arg  
 165 170 175  
 Phe Ala Gly Thr Pro Tyr Ala Thr Arg Ala Leu Val Asn Ala Arg Gly  
 180 185 190

210                      215                      220  
 Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp  
 225                      230                      235                      240  
 Ala Glu Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg  
 245                      250                      255  
 Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln  
 260                      265                      270  
 Leu Pro Gly Phe Asp Glu Gly Gly Gly Leu Arg Pro Xaa Lys  
 275                      280                      285

# 2 INFORMATION FOR SEQ ID NO:83:

1. SEQUENCE CHARACTERISTICS:
  - A) LENGTH: 173 amino acids
  - B) TYPE: amino acid
  - C) STRANDEDNESS: single
  - D) TOPOLOGY: linear

## 2. SEQUENCE DESCRIPTION: SEQ ID NO:83:

Thr Lys Phe His Ala Leu Met Gln Gln Gln Ile His Asn Glu Phe Thr  
 1                      5                      10                      15  
 Ala Ala Gln Gln Tyr Val Ala Ile Ala Val Tyr Phe Asp Ser Glu Asp  
 20                      25                      30  
 Leu Pro Gln Leu Ala Lys His Phe Tyr Ser Gln Ala Val Glu Glu Arg  
 35                      40                      45  
 Asn His Ala Met Met Leu Val His His Leu Leu Asp Arg Asp Leu Arg  
 50                      55                      60  
 Val His Ile Pro Gly Val Asp Thr Val Arg Asn Gln Phe Asp Arg Pro  
 65                      70                      75                      80  
 Glu Thr Val Lys Glu Leu Val Lys Val Glu Val Val Thr Val Thr Asp  
 85                      90                      95  
 His Val Gln Arg Leu Thr Ala Val Ala Arg Asn Glu Gly Asn Phe Leu  
 100                      105                      110  
 Glu His Gln Pro Met Gln Trp Pro Leu His Glu Gln Ile Glu Gln Val  
 115                      120                      125  
 Ala Leu Met Ala Thr Leu Val Arg Val Ala Asn Arg Ala Gly Ala Asn  
 130                      135                      140

165

170

# 2. INFORMATION FOR SEQ ID NO:84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 107 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:84:

```

Arg Ala Asp Glu Arg Lys Asn Thr Thr Met Lys Met Val Lys Ser Ile
1      5      10      15

Ala Ala Gly Leu Thr Ala Ala Ala Ala Ile Gly Ala Ala Ala Ala Gly
20      25      30

Val Thr Ser Ile Met Ala Gly Gly Pro Val Val Tyr Gln Met Gln Pro
35      40      45

Val Val Phe Gly Ala Pro Leu Pro Leu Asp Pro Xaa Ser Ala Pro Xaa
50      55      60

Val Pro Thr Ala Ala Gln Trp Thr Xaa Leu Leu Asn Xaa Leu Xaa Asp
65      70      75      80

Pro Asn Val Ser Phe Xaa Asn Lys Gly Ser Leu Val Glu Gly Gly Ile
85      90      95

Gly Gly Xaa Glu Gly Xaa Xaa Arg Arg Xaa Gln
100      105

```

# 3. INFORMATION FOR SEQ ID NO:85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```

Val Leu Ser Val Thr Glu Thr Asn Glu Leu Pro Val Ala Val Val Asn
1      5      10      15

Pro Leu Gly Gln Pro Ile Asp Gly Arg Gly Asp Val Asn Ser Asn Thr
20      25      30

Arg Arg Ala Leu Glu Leu Gln Ala Pro Ser Val Val Xaa Arg Gln Gly
35      40      45

```

```

65              70              75              80
Gly Lys Asn Arg Arg Leu Cys Arg Thr Pro Ser Ser Asn Gln Arg Glu
      85              90              95
Glu Leu Gly Val Arg Trp Ile Pro Arg Ser Arg Cys Ala Cys Val Tyr
      100             105             110
Val Gly His Arg Ala Arg Arg Gly Thr Tyr His Arg Arg
      115             120             125

```

2) INFORMATION FOR SEQ ID NO:86.

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 117 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:86:

```

Cys Asp Ala Val Met Gly Phe Leu Gly Gly Ala Gly Pro Leu Ala Val
1              5              10              15
Val Asp Gln Gln Leu Val Thr Arg Val Pro Gln Gly Trp Ser Phe Ala
      20              25              30
Gln Ala Ala Ala Val Pro Val Val Phe Leu Thr Ala Trp Tyr Gly Leu
      35              40              45
Ala Asp Leu Ala Glu Ile Lys Ala Gly Glu Ser Val Leu Ile His Ala
      50              55              60
Gly Thr Gly Gly Val Gly Met Ala Ala Val Gln Leu Ala Arg Gln Trp
      65              70              75
Gly Val Glu Val Phe Val Thr Ala Ser Arg Gly Lys Trp Asp Thr Leu
      80              85              90
Arg Ala Lys Lys Phe Ala Val Thr Thr Thr Thr Thr Thr Thr Thr
      95              100             105             110
Val Ser Ser Lys Gly
      115

```

3. INFORMATION FOR SEQ ID NO:87

1. SEQUENCE CHARACTERISTICS:
- A LENGTH: 117 amino acids
  - B TYPE: amino acid
  - C STRANDEDNESS: single

Met Tyr Arg Phe Ala Cys Arg Thr Leu Met Leu Ala Ala Cys Ile Leu  
 1 5 10 15  
 Ala Thr Gly Val Ala Gly Leu Gly Val Gly Ala Gln Ser Ala Ala Gln  
 20 25 30  
 Thr Ala Pro Val Pro Asp Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp  
 35 40 45  
 Pro Ala Trp Gly Pro Asn Trp Asp Pro Tyr Thr Cys His Asp Asp Phe  
 50 55 60  
 His Arg Asp Ser Asp Gly Pro Asp His Ser Arg Asp Tyr Pro Gly Pro  
 65 70 75 80  
 Ile Leu Glu Gly Pro Val Leu Asp Asp Pro Gly Ala Ala Pro Pro Pro  
 85 90 95  
 Pro Ala Ala Gly Gly Gly Ala  
 100

12 INFORMATION FOR SEQ ID NO:88:

12.1 SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 98 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

12.2 SEQUENCE DESCRIPTION: SEQ ID NO:88:

Met His Cys Arg Val Trp Leu Glu Ile Gln Trp Arg Gly Met Leu Gly  
 1 5 10 15  
 Ala Asp Glu Ala Arg Ala Gly Gly Ser Ala Arg Ile Trp Arg Glu His  
 20 25 30  
 Ser Met Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Ile Ala  
 35 40 45  
 Thr Lys Lys Thr Arg Lys Ile Lys Met Arg Val Trp Ser Ala Glu His  
 50 55 60  
 Glu Arg Lys Val Ala Glu Leu Thr Ser Asp Glu Ala Ala Ala Leu Gly  
 65 70 75 80  
 Asp Glu Leu Lys Gly Val Thr Ser  
 85

13 INFORMATION FOR SEQ ID NO:89:

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:89:

```

Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile
1           5           10           15
Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly
20           25           30
Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala
35           40           45
Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu
50           55           60
Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg
65           70           75           80
Ala Asp Glu Gln Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
85           90           95

```

(X2) INFORMATION FOR SEQ ID NO:90:

- 1. SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 166 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:90:

```

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
1           5           10           15
Arg Ala Asn Gln Val Gln Ala Pro Met Ala Asp Pro Pro Thr Asp Val
20           25           30
Ser Ile Thr Thr Thr Gln Leu Thr Thr Lys Lys Asn Ala Ala Gln Gln
35           40           45
Ala Val Ser Ser Ala Val Asn Met Ala Thr Thr Thr Thr Thr Thr Ala
50           55           60
Lys Gln Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Lys
65           70           75           80
Thr Gln Gln Val Asp Ala Gln Ala Ala Thr Ala Leu Asp Asn Asn Gly
85           90           95

```

Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp  
 130 135 140

Gln Gly Ala Ser Leu Ala His Xaa Gly Asp Gly Trp Asn Thr Xaa Thr  
 145 150 155 160

Leu Thr Leu Gln Gly Asp  
 165

## 2. INFORMATION FOR SEQ ID NO:91.

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

### xi. SEQUENCE DESCRIPTION: SEQ ID NO:91:

Arg Ala Glu Arg Met  
 1 5

## 3. INFORMATION FOR SEQ ID NO:92:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 263 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

### xi. SEQUENCE DESCRIPTION: SEQ ID NO:92:

Val Ala Thr Met Ser Val Thr Ala Gly Gln Ala Gln Leu Thr Ala Ala  
 1 10 15

Gln Val Arg Val Ala Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu Thr  
 20 25 30

Ala Pro Arg Ser Val Gln Ala Ala Asn Arg Ala Thr Leu Met Thr Leu  
 35 40 45

Ala Ala Thr Asn Val Leu Gly Ala Asn Thr Ser Ala Thr Ala Thr Asn  
 50 55 60

Gln Ala Gln Tyr Gln Gln Met Trp Ala His Asp Ala Ala Ala Met Pro  
 65 70 75 80

Gln Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Thr Trp Phe  
 85 90 95

```

115              120              125
Asn Asn Val Pro Gln Ala Leu Lys Gln Leu Ala Gln Pro Thr Gln Gly
130              135              140
Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser Pro
145              150              155              160
His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His Met
165              170              175
Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser Met
180              185              190
Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala Gln Ala Val Gln Thr Ala
195              200              205
Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu Gly
210              215              220
Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala
225              230              235              240
Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly
245              250              255
Arg Arg Asn Gly Gly Pro Ala
260

```

#### 4. INFORMATION FOR SEQ ID NO:93:

##### (i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 303 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

##### (ii) SEQUENCE DESCRIPTION: SEQ ID NO:93:

```

Met Thr Thr Ser Ser Ser Asn Glu His Tyr Ser Thr Ala Gln Pro Ala
1      5      10      15      20
Val Ser Thr Ser Gly Val Thr Ser Ser Thr Ala His Ala Asp Gly Gly
25     30     35     40     45     50
Ala Ser Lys Leu His Met Tyr Leu Asn Ile Ala Val Ala Val Leu Gly
55     60     65     70     75     80
Leu Ala Ala Tyr Ser Ala Ser Thr Glu Pro Met Pro Thr Leu Ser Thr
85     90     95     100    105    110

```

35	90	95
Pro Lys Ala Lys Ser His Val Thr Val Val Ala Val Leu Gly Val Leu		
100	105	110
Gly Val Phe Leu Met Val Ser Ala Thr Phe Asn Lys Pro Ser Ala Tyr		
115	120	125
Ser Thr Gly Trp Ala Leu Trp Val Val Leu Ala Phe Ile Val Phe Gln		
130	135	140
Ala Val Ala Ala Val Leu Ala Leu Leu Val Glu Thr Gly Ala Ile Thr		
145	150	155
Ala Pro Ala Pro Arg Pro Lys Phe Asp Pro Tyr Gly Gln Tyr Gly Arg		
165	170	175
Tyr Gly Gln Tyr Gly Gln Tyr Gly Val Gln Pro Gly Gly Tyr Tyr Gly		
180	185	190
Gln Gln Gly Ala Gln Gln Ala Ala Gly Leu Gln Ser Pro Gly Pro Gln		
195	200	205
Gln Ser Pro Gln Pro Pro Gly Tyr Gly Ser Gln Tyr Gly Gly Tyr Ser		
210	215	220
Ser Ser Pro Ser Gln Ser Gly Ser Gly Tyr Thr Ala Gln Pro Pro Ala		
225	230	235
Gln Pro Pro Ala Gln Ser Gly Ser Gln Gln Ser His Gln Gly Pro Ser		
245	250	255
Thr Pro Pro Thr Ile Phe Pro Ser Phe Ser Pro Pro Pro Pro Val Ser		
260	265	270
Ala Gly Thr Gly Ser Gln Ala Gly Ser Ala Pro Val Asn Tyr Ser Asn		
275	280	285
Pro Ser Gly Gly Gln Gln Ser Ser Ser Pro Gly Gly Ala Pro Val		
290	295	300

INFORMATION FOR SEQ ID NO 34

## 1. SEQUENCE CHARACTERISTICS

- A LENGTH: 307 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO: 34

```

GTCTTCGGGCG CGGCACTGCC GTTGGACCCG GCATCCGCCC CTGACSTCCC GACCGCCGCC      180
CAGTTGACCA GCCTGCTCAA CAGCCTGSCG GATCCCAACG TGTCGTTTGC GAACAAGGGG      240
AGTCTGGTCS AGGGCGGCAT CGGGGGCACC GAGGCGCGCA TCGCCGACCA CAAGCTGAAG      300
AAGGCCGCGG AGCACGGGGA TCTGCCGCTG TCGTTCAGCG TGACGAACAT CCAGCCGGCG      360
GCCGCCGGTT CGGCCACCGC CGACGTTTCC GTCTCGGGTC CGAAGCTCTC CTCGCCGGTC      420
ACGCAGAACC TCACGTTCTT GAATCAAGGC GGCTGGATGC TGTACCCCGC ATCGGCGATG      480
GAGTTGCTCC AGGCCGCGAG GAACTGA                                          507

```

2) INFORMATION FOR SEQ ID NO:95:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 168 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

2) SEQUENCE DESCRIPTION: SEQ ID NO:95:

```

Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala Ala Ala Ala
1      5      10      15
Ile Gly Ala Ala Ala Ala Gly Val Thr Ser Ile Met Ala Gly Gly Pro
20     25     30
Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro Leu Pro Leu
35     40     45
Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser
50     55     60
Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn Lys Gly
65     70     75     80
Gln Leu Val Val Lys Lys Lys Lys Thr Lys Ala Ala Lys Ala Asp
85     90     95
His Lys Leu Lys Gly Ala Ala Lys His His Asn Leu Pro Leu Lys Thr
100    105    110
Ser Val Thr Asn Ile Ser Ser Ala Ala Ala Thr Ser Ala Thr Ala Asp
115    120    125
Ala Ser Val Ser Lys Pro Lys Leu Ser Ser Lys Val Thr Lys Asn Val
130    135    140

```

Glu Leu Leu Gln Ala Ala Gly Asn  
165

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 500 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CGTGGCAATG TCSTTGACCG TCGGGGCCGG GGTGGGCTCC GCAGATCCCG TGGACGCGGT	60
CTTTAACACC ACCTGCAATT ACGGGCAGGT AGTAGCTGCC CTCAACCCGA CGGATCCGGG	120
GGCTCCCGCA CAGTTCAACG COTCACCAGT GGGCAGTCC TATTTGCCCA ATTTCTCTCC	180
TGCACCGCCA COTCAGCCCG CTGCCATGGC GGGCAATTG CAAGCTGTCC CCGGGGCGGC	240
AGAGTAGATG GGCCTTCTCG AGTCGGTTGC CGCTCTCTGC AAGAACTATT AAGCCCATGC	300
GGCCCCCATC GCGCGACCCG GCATCGTCCG CCGGGCTAGG CCAGATTGCC CCGCTCTCA	360
ACGGGCGCCA TCCCGCGACC CGGCATCGTC GCGGGGGGTA GGCCAGATTG CCGGCTCTCT	420
TACGGGGCGG CATCTCGTTC CGAATTCTTG CAGCCCGGGG GATCCACTAG TTCTAGAGCG	480
GGCGGCACCG GGGTGGAGGT	500

(3) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 96 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Val Ala Met Ser Leu Thr Val Val Ala Gly Thr Ala Ser Val Asp Pro	10	20	30
Val Asp Ala Val Ile Asn Thr Thr Lys Asn Tyr Gly Gln Val Val Ala	40	50	60
Val Leu Asn Ala Thr Asp Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser	70	80	90
Pro Val Ala Gln Ser Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr	100	110	120

Gln Tyr Ile Gly Leu Val Glu Ser Val Ala Gly Ser Cys Asn Asn Tyr  
                   85                                  90                                  95

(2) INFORMATION FOR SEQ ID NO:98:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 154 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:98:

ATGACAGAGC AGCAGTGGAA TTTCGCGGGT ATCGAGGCCG CGGCAAGCGC AATCCAGGGA 60  
 AATGTACAGT CCATTGATTC GGTGCTGAC GAGGGGAAGC AGTCCCTGAC TAAGCTGGCA 120  
 GCGGCTGGG GCGGTAGCGG TTGGAAGCG TACC 154

(2) INFORMATION FOR SEQ ID NO:99:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 51 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:99:

Met Thr Glu Glu Ile Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser  
 1                                  10                                  15  
 Ala Ile Glu Glu Asn Val Thr Ser Ile His Ser Leu Leu Asn Glu Gly  
 21                                  25                                  30  
 Lys Glu Ser Leu Thr Tyr Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser  
 35                                  40                                  45  
 Ala Ala Tyr  
 51

(2) INFORMATION FOR SEQ ID NO:100:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 282 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GCTCGAAACG CGGCACAGCC GACGGTGGCT CCGNCAGAGC GCTGCTGCA AAATCGTTGA 180  
 GACAATTCCN CGGGGGGCCC TACAAGGAAG TCGGTGCTGA ATTGNCNGG TATCTGCTCG 240  
 ACCTGTGTGG TCTGNAGCCG GACGAAGCGG TGCTCGACGT CG 282

(2) INFORMATION FOR SEQ ID NO:101:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:101.

GATGCTACCC GTGCGAGTGC TCGGGCCCTT TGAGGATGGA GTGCACCTGT CTTCCTGAT 50  
 GGCATACCCA GAGATGTTGG CGGCGGCGCC TGACACCTG CAGAGCATCG GTGCTACCA 120  
 TGTGCTTACG AATGCGCTG CGGCGGCGCC GAGGACTGCG GTGCTGCGCC CGCTGCGCA 180  
 TGAGGTGTGG GCGCTGACTG CGGCGCACTT CGCGGCACAT CGCGCGATGT ATCAGTCTGT 240  
 GAGCGCTCGG GCTGCTGCGA TTCATGACCA GTTCGTGCGC ACCCTTCCCA GCAGCGCCAG 300  
 GTGCTATGCG GCGCTGAAG TCGCCAATGC GCGCGCGCGC AGCTAAGCCA GGAACAGTCC 360  
 GCACGAGAAA CCACGAGAAA TAGGGACACG TAATGCTGGA TTTCGGGCGC TTACACCGCC 420  
 AGATGAATCG CGCGAGGATG TACCGCGCGC CGGCTTCCGC TTGCTGCTG GCGCGCGCTG 480  
 AATGCTGCGA GAGCTGCGCG ATGAGCTGT GTTCGCGCGC TTGCGCTTT GAGTGGTGG 540  
 GTGCGGTGT GACGCTGCGC TCGTGATAG GTTCGTGCGC GGTGTGATG GTGCGCGCGC 600  
 GTCGCGCTA TGTGGCTGG ATGAGCTGA TCGCGCGCGC GCGCGAGCTG ACCGCGCGCT 660  
 GGTGCGCTT TGTGCGCGC GATGAGGA GATGAGTA GTGAGCTT GGTGCGCGC 720  
 GATGAGGA GAGCTGCTT GAGTGAATG GTGAGTA GTGAGCTT GGTGCGCGC 780  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 840  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 900  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 960  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1020  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1080  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1140  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1200  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1260  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1320  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1380  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1440  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1500  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1560  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1620  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1680  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1740  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1800  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1860  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1920  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 1980  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2040  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2100  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2160  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2220  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2280  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2340  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2400  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2460  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2520  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2580  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2640  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2700  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2760  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2820  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2880  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 2940  
 GATGAGCTT GATGAGCTT GAGTGAATG GATGAGTA GTGAGCTT GGTGCGCGC 3000

CGATGACCAA CTCGGGTGTG TCGATGACCA ACACCTTGAG CTCGATGTTG AAGGGCTTTG	1200
CTCCGCGCGC GCGCGCCAG GCCGTGCAAA CCGCGGCGCA AAACGGGGTC CGGGCGATGA	1260
JCTCGGTGGG CAGCTGCTG GTTCTTTGGG GTCTGGGCGG TGGGGTGGCC GCCAACTTGG	1320
GTGGGGCGGC CTCGGTGGGT TCGTTGTGGG TCCCGCAGGC CTGGGCGCGC GCCAACCAGG	1380
CAGTCACCCC GCGCGGCGCG GCGCTGCGGC TGACCAGGCT GACCAGCGCC GCGGAAAGAG	1440
GGCCCGGGCA GATGCTGGGC GGGCTGCCGG TGGGGCAGAT GGGCGCCAGG GCCGGTGGTG	1500
GGGTCACTGG TGTGCTGGT GTTCCGCGCG GACCTATGT GATGCCGCAT TCTCGGCGCG	1560
CGCGCTAGGA GAGGGGCGCG AGACTGTCTT TATTTGACCA GTGATCGGCG GTCTCGGTGT	1620
TTCCCGCGCG CGCTATGACA ACAGTCAATG TGCATGACAA GTTACAGGTA TTAGGTCCAG	1680
TTTCAACAAG GAGACAGGCA ACATGGCGTC AGTTTTATG ACCGATCCCG ACCCGATCGC	1740
GGACATGGCG GCGCGTTTTG AGGTGCACGC CGAGACGGTG GAGGACGAGG CTCGCGCGAT	1800
GTGGCGCTCC GCGCAAAACA TTTCCGCTCC GGGGTGGAGT GGGATGGCGG AGGCGACCTC	1860
GCTAGACACC ATGGCGCAGA TGAATCAGCG GTTTCGCAAC ATGCTGAACA TGCTGCACGG	1920
GCTCGCTGAC CGGCTGGTTG GCGACGCCAA CAACTACGAG TAGCAAGAGC AGGCTTCCCA	1980
GCACATGCTC AGCAGCTAAC CTCAGCGGCT GCAGGACAAT ACTTTTACAA GCGAAGGAGA	2040
ACAGGTTTGA TGACGCTCAA CTATCAATTC GGGGATGTCT ACCGTACCGG CGCCATGATC	2100
CGCGCTGAGG CGGGGTTGCT CGAGCGCGAG CATCAGGCCA TCATTGCTGA TGTGTTGAGC	2160
CGAGTGAAT TTGCGCGCGG CGCGGCTTCC CGGCGCTGCC AGGGGTTGAT TACCGAGTTC	2220
CGAGTGAAT CGAGTGAAT CTACGACGAG CGCAACGCGC ACAGGACAGG CGTGCAGGCT	2280
CGAGTGAAT ACAGTGAAT ACAGTGAAT CGAGTGAAT CGAGTGAAT CGAGTGAAT	2340
CGAGTGAAT ACAGTGAAT ACAGTGAAT CGAGTGAAT CGAGTGAAT CGAGTGAAT	2400
CGAGTGAAT ACAGTGAAT ACAGTGAAT CGAGTGAAT CGAGTGAAT CGAGTGAAT	2460
CGAGTGAAT ACAGTGAAT ACAGTGAAT CGAGTGAAT CGAGTGAAT CGAGTGAAT	2520
CGAGTGAAT ACAGTGAAT ACAGTGAAT CGAGTGAAT CGAGTGAAT CGAGTGAAT	2580
CGAGTGAAT ACAGTGAAT ACAGTGAAT CGAGTGAAT CGAGTGAAT CGAGTGAAT	2640
CGAGTGAAT ACAGTGAAT ACAGTGAAT CGAGTGAAT CGAGTGAAT CGAGTGAAT	2700

```

GGTGGCCACC CACGGCCAGG AGGGCTTCGG GGTGGCTGCC ATCAGATTGG CTGGCTAGTG      2880
GGTTCTGCAG CGGTGCCAGG CCGCTGCCGG CAGGGTGGCG CCGATCGCGG CCACCAGGCC      2940
GGCGTGGGCG TCGTGGTGA CCAGCCCGAC CCGGACAGG CCGCGGGCGA CCAGGTCCGC      3000
GAAGAACGCC AGCCAGCCGG CCGCGTCTC GCGGAGGTG ACCTGGATGC CCAGGATC      3058

```

## (2) INFORMATION FOR SEQ ID NO:102:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 391 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:102:

```

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met      1
1           5           10           15
Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp      20
20          25          30
Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser      35
35          40          45
Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly      50
50          55          60
Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr      65
65          70          75
Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala      80
80          85          90          95
Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala      100
100         105         110
Asn Asn Arg Ala Ile Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Ile      115
115         120         125
Gln Asn Thr Pro Ala Ile Ala Val Asn Ile Ala Gln Thr Ile Thr Met      130
130         135         140
Trp Ala Gln Asp Ala Ala Ala Met Pro Gly Tyr Ala Ala Ala Thr Ala      145
145         150         155         160
Thr Ala Thr Ala Thr Leu Leu Pro Ile Ile Gln Ala Pro Ile Met Thr      165
165         170         175

```

Gln Gln Leu Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu  
210 215 220

Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn  
225                   230                   235                   240

Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val  
245 250 255

Ser Met Thr Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala  
260 265 270

Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala  
375 380 385

Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly  
290 295 300

Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val  
305 310 315 320

Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg  
325 330 335

Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly  
340 345 350

Phe Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly  
155                      160                      165

17.   Leu Ser Gly Val   Leu Arg Val   Pro Pro Arg Pro Tyr Val Met  
          100                   100                   100

Pro His Ser Pro Ala Ala Gly  
185 390

[illegible]

### INCIDENT CHARACTERISTICS

- ```

A LENGTH 1125 base pairs
B TYPE nucleic acid
C STRANDEDNESS single
D TOPOLOGY linear

```

4. SEQUENCE DESCRIPTION. SEQ ID NO: 103

ALL INFORMATION CONTAINED HEREIN IS UNCLASSIFIED

CCGTGTGGGG TCGATTGGG CGGACCAGTC CTCACCAACG CTTGGCGTGC GCGCCAGGCG 240  
 GCGGATCAGA TCGCTTGACT ACCAATCAAT CTTGAGCTCC GGGGCGGATG CTCGGGCTAA 300  
 ATGAGGAGGA GCACCGCTGT CTTTCACTGC GCAACCGGAG ATGTTGGCGG CCGCGGCTGG 360  
 CGAACTTCGT TCCCTGGGGG CAACGCTGAA CGCTAGCAAT GCGGCGGAG CCGTGCGGAC 420  
 GACTGCGGTG CTGCCCCCGG CTGCGGACGA GGTGTCGCTG CTGCTTGCCA CACAATTCCG 480  
 TACGCATGCG GCGACGTATC AGACGGCCAG CGCCAAGGCC GCGGTGATCC ATGAGCAGTT 540  
 TGTGACCAGC CTGCGCACCA GCGCTAGTTC ATATCGGGAC ACCGAGGCGG CCAACCGCTGT 600  
 GGTGACCGGC TAGCTGACCT GACCGTATTC GAGCGGAAGG ATTATCGAAG TGGTGGATTT 660  
 GGGGGGCTTA GCACTGAGGA TCACTCCCG GAGGATGTAC GCGGCGGCGG GTTGGGCTTC 720  
 GCTGCTGGCC GCGCGGAGA TGTGGGACAG TGTGGCGAGT GACCTGTTTT GCGCGCGCTC 780  
 GCGCTTTTCAG TCGGTGCTGT GGGGTCTGAC GCTGCGGCTC TCGATAGGTT TGTGCGCGGG 840  
 TGTGATGGCG GCGGCGGCTT CGCGTATGT GCGCTGGATG AGCGTCACCG GGGGCGAGGC 900  
 CGAGCTGACC GCGCGCGAGG TCGGGGTTGC TCGCGCGGCC TACGAGACAG CGTATAGGCT 960  
 GACGGTSCCG CCGCGCGTGA TCGCGGAGAA CGCTACCGAA CTGATGACCG TGACCGCGAC 1020  
 GACCTCTTTC GGGGAAAAA CCGCGCGGAT CGAGCGCAAT CAGGCGCGCAT ACAGCGCAGT 1080  
 GTGCGCGGCA GACCGCGAGG CGATGTATCG CTAGCGCGCG ACGGCGCGCG CCGCGACGGA 1140  
 GCGCTTCTTC CGCTTTCAGG AGCGCGCACT GATGACGAGC TCGCGCGCGG TCGTTGAGCA 1200  
 GCGCTTCTTC GTGACGAGG CGATGACAGC GCGCGCGCGG AACCACTTGA TGAACATGT 1260  
 GCGCTTCTTC GTGACGAGG TCGCGCGAGG AGCGCGCGG GTGCTACCTT GTTCAAGCT 1320  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1380  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1440  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1500  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1560  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1620  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1680  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1740  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1800  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1860  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1920  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 1980  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2040  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2100  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2160  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2220  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2280  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2340  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2400  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2460  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2520  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2580  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2640  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2700  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2760  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2820  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2880  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 2940  
 GCGCTTCTTC TCGACGCGCG TCGCGCGGCA TCGCTTCTTC GTGACGAGC TCGCTTCTTC 3000

(A) LENGTH: 159 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:104:

```

Val Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
1             5             10             15
Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Lys Met Trp
20             25             30
Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
35             40             45
Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
50             55             60
Leu Met Ala Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
65             70             75             80
Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
85             90             95
Ala Tyr Glu Thr Ala Tyr Arg Leu Thr Val Pro Pro Pro Val Ile Ala
100            105            110
Gln Asn Arg Thr Gln Leu Met Thr Leu Thr Ala Thr Asn Leu Leu Gly
115            120            125
Gln Asn Thr Pro Ala Ile Glu Ala Asn Gln Ala Ala Tyr Ser Gln Met
130            135            140
Pro Gly Gln Asp Ala Glu Ala Met Tyr Gly Tyr Ala Ala Thr Ala Ala
145            150            155            160
Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr
165            170            175
Asn Pro Gly Gln Leu Leu Thr Ala Ala Val Ala Val Gln Ile Ala Ile
180            185            190
Asp Thr Ala Ala Ala Asn Thr Leu Met Asn Asn Val Pro Gln Ala Leu
195            200            205
Gln Gln Leu Ala Gln Pro Ala Gln Gly Val Val Pro Ser Ser Lys Leu
210            215            220
Gln Thr Leu Thr Thr Ala Val Ser Gln Thr Leu Ser Pro Leu Ser Asn
225            230

```

Ser Met Thr Asn Thr Leu His Ser Met Leu Lys Gly Leu Ala Pro Ala  
 260 265 270

Ala Ala Gln Ala Val Glu Thr Ala Ala Glu Asn Gly Val Trp Ala Met  
 275 280 285

Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu  
 290 295 300

Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser  
 305 310 315 320

Leu Ser Val Pro Pro Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro  
 325 330 335

Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr  
 340 345 350

Ala Pro Gly His Met Leu Gly  
 355

# 1. INFORMATION FOR SEQ ID NO:105:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3027 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (E) SEQUENCE DESCRIPTION: SEQ ID NO:105:

ATTTCAGCTG AGAATGATATG TGAAGGCGCTG TATCCACCGAT GGCTGAGACA ACCGAACCAAT 60  
 TGTCCGACCG GGGGACATCG GAAGGCGAGC CGATGCGCTT GGGGCGCGAA GCGGAAGCGG 120  
 GCGAAGCGCA AGCGCTGGCG GCGCGCGCGG GGGCGCGTGC CGCTCGCGCG CGCTTGAAGC 180  
 CTCAGGCGCT GCGGATGGCT CAGGCGGAGG ACCGAAACCT GCGCGAGGAT ATCCAGACTG 240  
 GAAAGAGTTC GAAGACTATG AGAATATATG GAATATGAG TGTGAGAG AGGAGGCGG 300  
 GATTCGCGCA TGTTCGCGA GTTCCTTCTG GTTCGCTTTC GAGAGATTC CAGGATTTC 360  
 ATTCGCTGTC GAATGCTTTC TATCTTCTG GTTCGCGG GTCAGCGGAT ACATTCTGTC 420  
 GAAAGCGCT GAGCGCAATG AACGCGAGCA GCGGCGCG GCGTTCGCG GCGGAGCGAA 480  
 GGAAGCTGTC ATCAACATGA GCTGCGTGGG GTTCGAGAGG GCGAAGAGAG AGCTCGCGCG 540  
 GATGATGAG AGCTCGAG GCGAATTCAG CGATGATTCG GAGGAGGAG GAGCGGATTC 600

|                                                                   |      |
|-------------------------------------------------------------------|------|
| TTCCGCTGGG GCGAAAGACG AACCACTGC GTGGCGGCTC AAAGTGACCG TGACCGAAGA  | 780  |
| GGGGGGACAG TACAAGATGT CGAAAGTTGA GTTCGTACCG TGACCGATGA CTACCGGAC  | 840  |
| GTCAACACCG AAACCACTGA CGCCACCGAA GTCGCTGAGA TCGACTCAGC CGCAGGCGAA | 900  |
| GCGGGTGATT CGGCGACCGA GGCATTTGAC ACCGACTCTG CAACGGAATC TACCGCCAC  | 960  |
| AAGGGTCAGC GGCACCGTGA CCTGTGGCGA ATGCAGGTTA CTTGAAACC CTTCCGGTG   | 1020 |
| ATTCTCATCC TGCTCATGTT GATCTCTGGG GGCGCGACCG GATGGCTATA CTTGAGCAA  | 1080 |
| TACGACCGCA TCAGCAGACG GACTCCGGCG CGCGCCGTGC TGCCCTCGCC GCGGCGTCTC | 1140 |
| ACGGGACAAT CGGCTGTTG TGTATTGACC CGACACGTCG ACCAAGACTT CGCTACCGCC  | 1200 |
| AGGTCCGACC TCGCGCGCGA TTTGCTGTCC TATACGACCA GTTCAGCGAG CAGATCCTGG | 1260 |
| GTCCGGCGGC CAAACAGAAG TCACTGAAAA CGACCGCGAA GTGCTGCGC CGCGCCGTCT  | 1320 |
| CGGAGCTACA TCCGGATTCC GCGCTGCTTC TGGTTTTTGT CGACGAGAGC ACTACGAGTA | 1380 |
| AGGACAGCGC CAATCCGTCG ATGGCGGCGA CGAGCGTGAT GGTGACCTTA GCGAAGCTCG | 1440 |
| ACGGCAATTG GGTGATCACC AAGTTCACCC CGGTTTAGGT TGCGGTAGGC GGTGCGCAAG | 1500 |
| TCTGACCGGC GCGCGGGTGG CTGCTGCTGC GAGATACCGC CGTTCTGCG GACAATCAGC  | 1560 |
| CGCCGACCTG AAACAGATCT CGGCGGCTGT CTAATCGGCG GGGTTATTTA AGATTAGTTG | 1620 |
| CGATGCTATT TACCTGATCT TCAGATTGTT CAGCTGGATT TAGCTTCGCG CGAGGGCGGC | 1680 |
| TGCTGCACTT TGCATCTCGG GTTGTGACTA GTTCAGAGAA TTGACCTGT TCGGACCTT   | 1740 |
| GTTCGCTGTC CATGATTGCT CCTAGTTATG CGCGAGCGCA AGGATTATCG AAGTGTGGA  | 1800 |
| ATTGCGGCGC TTACGACCGG AGATCAACTC CGCGAGGATG TACCGCGCGC CGGCTTCGCG | 1860 |
| TTGCTGCTG TCGCGCGCGA AGATGTGCGA CAGCTGCGCG AGTGAATCTT TTGCGCGCTT  | 1920 |
| TTGCTGCTG TCGCGCGCGA AGATGTGCGA CAGCTGCGCG AGTGAATCTT TTGCGCGCTT  | 1980 |
| TTGCTGCTG TCGCGCGCGA AGATGTGCGA CAGCTGCGCG AGTGAATCTT TTGCGCGCTT  | 2040 |
| TTGCTGCTG TCGCGCGCGA AGATGTGCGA CAGCTGCGCG AGTGAATCTT TTGCGCGCTT  | 2100 |
| TTGCTGCTG TCGCGCGCGA AGATGTGCGA CAGCTGCGCG AGTGAATCTT TTGCGCGCTT  | 2160 |
| TTGCTGCTG TCGCGCGCGA AGATGTGCGA CAGCTGCGCG AGTGAATCTT TTGCGCGCTT  | 2220 |
| TTGCTGCTG TCGCGCGCGA AGATGTGCGA CAGCTGCGCG AGTGAATCTT TTGCGCGCTT  | 2280 |

```

TGTGCCCCAA GCGCTGCAAC AACTGGCCCA GCGCACGAAA AGCATCTGGC CGTTCGACCA   2460
ACTGAGTGAA CTCTGGAAGG CCATCTCGCC CCATCTGTCC CCGCTCAGCA ACATCGTGTC   2520
GATGCTCAAC AACCACGTGT CGATGACCAA CTCGGGTGTG TCGATGGCCA GCACCTTGCA   2580
CTCAATGTTG AAGGGCTTTG CTCGGGCGGC GGCTCAGGCC GTGGAACCG CGGCGCAAAA   2640
CGGGGTCCAG GCGATGAGCT CGCTGGGCAG CCAGCTGGGT TCGTCGCTCG GTTCTTCGGG   2700
TCTGGGCGCT GGGGTGGCCG CCAACTTGGG TCGGGCGGCC TCGGTGCGTT CGTTGTGCGT   2760
GCGCCAGGCC TGGGCGCGCG CCAACCAGGC CGTCACCCCG GCGGCGCGGG CGCTGCCGCT   2820
GACCAGCGTG ACCAGCGCCG CCAAAACCGC CCGCGGACAC ATGCTGGGCG GGCTACCGCT   2880
CGCGCAACTG ACCAATAGCG CCGCGCGGTT TGGCGGGGTT AGCAATCGCT TCGCGATGCC   2940
CGCGCGGGCG TACGTAATCG CCGGTGTGCC TCGCGCGGGG TAACGCGGAT CGGCACGCAA   3000
TCGCGGCGCT CTATCGCGCG ACGGATC                                     3027

```

## 2. INFORMATION FOR SEQ ID NO:106:

### 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 396 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

### 2. SEQUENCE DESCRIPTION: SEQ ID NO:106:

```

1  Val Asp Phe Ile Ala Met Trp Arg Ile Ile Asn Ser Ala Arg Met
15
Tyr Ala Gly Pro Ile Ser Ala Ser Leu Val Ala Ala Ala Lys Met Trp
25
Ser Ser Met Ala Ser Ser Ser Ser Ser Ala Ala Ser Ala Phe Ile Ser
35
Ala Val Thr Phe Ser Ser Ser Ile Ser Thr Ile Gly Ser Ser Ala Gly
45
Leu Met Val Ala Ala Ala Ser Arg Thr Val Ala Trp Met Ser Val Thr
55
Ala Ile Ile Ala Ile Ser Thr Ala Ala Ile Val Arg Val Ala Ala Ala
65

```

|                                                                 |     |         |
|-----------------------------------------------------------------|-----|---------|
| 115                                                             | 120 | 125     |
| Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met |     |         |
| 130                                                             | 135 | 140     |
| Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Thr Ala Ala |     |         |
| 145                                                             | 150 | 155 160 |
| Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr |     |         |
| 165                                                             | 170 | 175     |
| Asn Pro Gly Gly Leu Leu Glu Gln Ala Val Ala Val Glu Glu Ala Ile |     |         |
| 180                                                             | 185 | 190     |
| Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu |     |         |
| 195                                                             | 200 | 205     |
| Gln Gln Leu Ala Gln Pro Thr Lys Ser Ile Trp Pro Phe Asp Gln Leu |     |         |
| 210                                                             | 215 | 220     |
| Ser Glu Leu Trp Lys Ala Ile Ser Pro His Leu Ser Pro Leu Ser Asn |     |         |
| 225                                                             | 230 | 235 240 |
| Ile Val Ser Met Leu Asn Asn His Val Ser Met Thr Asn Ser Gly Val |     |         |
| 245                                                             | 250 | 255     |
| Ser Met Ala Ser Thr Leu His Ser Met Leu Lys Gly Phe Ala Pro Ala |     |         |
| 260                                                             | 265 | 270     |
| Ala Ala Gln Ala Val Glu Thr Ala Ala Gln Asn Gly Val Gln Ala Met |     |         |
| 275                                                             | 280 | 285     |
| Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu |     |         |
| 290                                                             | 295 | 300     |
| Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser |     |         |
| 305                                                             | 310 | 315 320 |
| Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro |     |         |
| 325                                                             | 330 | 335     |
| Ala Ala Asn Ala Leu Thr Leu Thr Ser Leu Thr Ser Ala Val Gln Thr |     |         |
| 340                                                             | 345 | 350     |
| Ala Pro Gly His Met Leu Gly Gly Leu Pro Leu Gly Gln Leu Thr Asn |     |         |
| 355                                                             | 360 | 365     |
| Ser Gly Gly Gly Phe Gly Gly Val Ser Asn Ala Leu Arg Met Pro Pro |     |         |
| 370                                                             | 375 | 380     |
| Leu Ala Thr Val Met Ser Asn Val Pro Ala Ala Ile                 |     |         |

## (i) SEQUENCE CHARACTERISTICS.

- (A) LENGTH: 1616 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

```

CATCGGAGGG AGTGATCACC ATGCTGTGGC ACGCAATGCC ACCGGAGTAA ATACCGCACG      60
GCTGATGGCC GCGCGGGGTC CGGCTCCAAT GCTTGCGGCG GCGCGGGAT GGCAGACGCT      120
TTGCGCGGCT CTGGACGCTC AGGCCGTCCA GTTGACCGCG CGCCTGAACT CTCTGGGAGA      180
AGCCTGGACT GGAGGTGGCA GCGACAAGCC GCTTGCGGCT GCAACGCCCA TGGTGGTCTG      240
GCTACAAACC GCGTCAACAC AGGCCAAGAC GCGTGGATG CAGGCGAGCC GCGAAGCCCG      300
GCGATACACT CAGGCGATGG GACGAGCCCG CTCGCTGCGG GAGATCGCCG CCAACCCACAT      360
CATCCAGGCG GTGCTTACCG CCACCAACTT CTTCGGTATC AACACGATCC CGATCGCGTT      420
GACCGAGATG GATTATTTCA TCGGTATGTG GAACGAGGCA GCGCTGGCAA TGGAGGTCTA      480
CGAGGCGGAG ACCGCGGTTA ACACGCTTTT CGAGAAGCTC GAGCGGATGG COTCGATCCT      540
TGATCGCGGC GCGAGCCAGA GCACGACGAA CCGGATCTTC GGAATGCGCT CCGCTGGCAG      600
CTCAACACCG GTTGGCGAST TCGCGCGGCG GCTTACCGAG ACCTGCGGCG AACTGGGTGA      660
GATGACCGCG CGCATCGAGC AGCTGACGCA TCGGCTGAG CAGGTGACCT GCTTGTTCAG      720
TGAGGTGGCG GCGACCGCGC GCGGCAACCG AGTTGACGAG GAAGCGCGCG AGATGGGCTT      780
GCTCGGCAAC AGTCCGCTGT CGAAGCATCC GTTGGTGGT CGATCAGGCG CGAGCGCGCG      840
CGCGCGCGTG CTGCGCGCGG AGTCCGTACC TCGGCGAGGT GGTGCGTTGA CGCGACCGCG      900
AGTGAAGTGT CAGGTGATCG AAAAGCGCGT TCGGCGTGG CTGATGCGCG GAGTGTGTG      960
GAGAGTGTG CCAAGCGT TCGCGCGT GTTGTGCGT GAGTGTGTG GAGTGTGTG      1020
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1080
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1140
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1200
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1260
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1320
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1380
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1440
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1500
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1560
GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG GAGTGTGTG      1616

```

```

AATATTGCTC AGGCCGGCGT CCAATACTCG AGGGCCGACG AGGAGCAGCA GCAGGGCGCTG      1500
TCCTCGCAAA TGGGCTTCTG ACCCGCTAAT ACGAAAAGAA ACGGAGCAAA AACATGACAG      1560
AGCAGCAGTG GAATTTTCGGG GGTATCGAGG CCGCGGCAAG CGCAATCCAG GGAAT      1616

```

## (2) INFORMATION FOR SEQ ID NO:108:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 432 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:108:

```

CTASTGGATG GGACCATGGC CATTTTCTGC AGTCTCACTG CTTCTGTGTG TGACATTTTG      60
GCACGCGCGC GGAAACGAAG CACTGGGGTC GAAGAACGGC TGCCTGCCA TATCGTCCGG      120
AGTTTCATA CTTCTGTGCG GCGGAAGAG CTTGTCTAG TCGCGCGCA TGACAACCTC      180
TCAGAGTGGC CTCAAACGTA TAAACACGAG AAAGGCGAG ACCACGGAA GGTGGAATC      240
GCGGATCCC CTGTTTGGCT ATTCTACGCG AACTCGGCGT TCGCTATGC GAACATCCCA      300
GTGAGCTTGC CTTGGGTGCA AGCCATTGCC TGACCGGCTT CGCTGATCCT CCGGCGCAGG      360
TTCTGAGCG CTTGTTGAG CTGGGTAGCC GTGGCTGCC ATTTTTCCTG GACAGCCTCG      420
TAAAGCTGCG AA      480

```

## (iii) INFORMATION FOR SEQ ID NO:109:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 368 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:109:

```

Met Leu Trp His Arg Met Ser Pro Glu Ala Asn Thr Val Arg Ser Met
1          20          30          40          50
Met Gly Ala Gly Pro Ala Pro Met Leu Ala Ala Ala Ala Gly Trp Glu
1          20          30          40          50
Glu Leu Ser Ala Ala Ser Thr Ala Thr Ala Thr Val Thr Thr Ala Arg
1          20          30          40          50

```

Leu Ala Ala Ala Thr Pro Met Val Val Trp Leu Gln Thr Ala Ser Thr  
 65 70 75 80  
 Gln Ala Lys Thr Arg Ala Met Gln Ala Thr Ala Gln Ala Ala Ala Tyr  
 85 90 95  
 Thr Gln Ala Met Ala Thr Thr Pro Ser Leu Pro Glu Ile Ala Ala Asn  
 100 105 110  
 His Ile Thr Gln Ala Val Leu Thr Ala Thr Asn Phe Phe Gly Ile Asn  
 115 120 125  
 Thr Ile Pro Ile Ala Leu Thr Glu Met Asp Tyr Phe Ile Arg Met Trp  
 130 135 140  
 Asn Gln Ala Ala Leu Ala Met Glu Val Tyr Gln Ala Glu Thr Ala Val  
 145 150 155 160  
 Asn Thr Leu Phe Glu Lys Leu Glu Pro Met Ala Ser Ile Leu Asp Pro  
 165 170 175  
 Gly Ala Ser Gln Ser Thr Thr Asn Pro Ile Phe Gly Met Pro Ser Pro  
 180 185 190  
 Gly Ser Ser Thr Pro Val Gly Gln Leu Pro Pro Ala Ala Thr Gln Thr  
 195 200 205  
 Leu Gly Gln Leu Gly Glu Met Ser Gly Pro Met Gln Gln Leu Thr Gln  
 210 215 220  
 Pro Leu Gln Gln Val Thr Ser Leu Phe Ser Gln Val Gly Gly Thr Gly  
 225 230 235 240  
 Thr Gly Asn Pro Ala Asp Glu Gln Ala Ala Gln Met Gl Leu Leu Gly  
 245 250 255  
 Thr Ser Pro Leu Ser Asn His Pro Leu Ala Gly Gly Ser Gly Pro Ser  
 260 265 270  
 Ala Gly Ala Gln Leu Leu Arg Ala His Ser Leu Pro Gly Thr Gly Ser  
 275 280 285  
 Ser Leu Thr Ala Thr Pro Leu Met Ser His Leu Ile Glu Gly Arg Val  
 290 295 300  
 Ala Pro Ser Val Met Pro Ala Ala Ala Ala Gly Ser Ser Ala Thr Gly  
 305 310 315 320  
 Thr Ala Ala Pro Val Gly Ala Gly Ala Met Gly Gln Gln Ala Thr Ser  
 325 330 335

355

360

365

## 2. INFORMATION FOR SEQ ID NO:110:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 100 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (2) SEQUENCE DESCRIPTION: SEQ ID NO:110:

```

Met Ala Glu Met Lys Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly
1           5           10           15

Asn Phe Glu Arg Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val
20           25           30

Glu Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly
35           40           45

Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys
50           55           60

Gln Lys Gln Glu Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly
65           70           75           80

Val Gln Tyr Ser Arg Ala Asp Glu Gln Gln Gln Gln Ala Leu Ser Ser
85           90           95

Gln Met Gly Phe
100

```

## 3. INFORMATION FOR SEQ ID NO:111:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 196 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (2) SEQUENCE DESCRIPTION: SEQ ID NO:111:

```

GATGTCGGGG GAGCTGAAAA GCGAGATCGA GCGCTGGAG TCGAGGCGAG GTTCTTTGCA      60
GGGCACTGG GCGGGGGGGG GCGGGAGCGG GCGGAGGGG GCGCTGTCTG GTTTCGAGA      120
AGTACCTAAT AATGAGAAGG AGGAACTGGA GAGATCTCG ACGAATATT GTGAGGCGGG      180

```

CTTGACGAGG GGAAGCAGTC COTGACCAAG CTCGCA

396

(2) INFORMATION FOR SEQ ID NO:112:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 80 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

```

Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala
1           5           10           15

Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln
20           25           30

Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu
35           40           45

Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser
50           55           60

Arg Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
65           70           75           80
  
```

(2) INFORMATION FOR SEQ ID NO:113:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 387 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

```

TAAATGCGA ATGAGTGTCT TGGTATTCT AATGAACTT AGGCTTCTT TATTTGAACTA
101

TAAATGCGA ATGAGTGTCT TGGTATTCT AATGAACTT AGGCTTCTT TATTTGAACTA
102

TAAATGCGA ATGAGTGTCT TGGTATTCT AATGAACTT AGGCTTCTT TATTTGAACTA
103

ATGATGACCG ATGAGTGTCT TGGTATTCT AATGAACTT AGGCTTCTT TATTTGAACTA
104

TAAATGCGA ATGAGTGTCT TGGTATTCT AATGAACTT AGGCTTCTT TATTTGAACTA
105

TAAATGCGA ATGAGTGTCT TGGTATTCT AATGAACTT AGGCTTCTT TATTTGAACTA
106
  
```

## 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 272 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:114:

```

CGGCACGAGG ATCTCGGTTG CCCCAACGGC GCTGGCGAGG GCTCCGTTCC GGGGGCGAGC      60
TGCSCGCCCGG ATGCTTCCTC TGCCCGCAGC CGCGCCTGGA TGGATGGACC AGTTGCTACC      120
TTCCCGACST TTCGTTGGGT GTCTGTCCGA TAGCGGTGAC CCCGGCGCGC ACGTGGGGAG      180
TGTGGGGGGG CAGGCGCGGT CGGTGCTTCG CCCGGGGAGC CAGACGGTCT GGACGGAACS      240
CGCGGGGGTT CGCCGATTGG CATCTTTGCC CA      300

```

## 2. INFORMATION FOR SEQ ID NO:115:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:115:

```

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val
1           5           10           15
Val Ala Ala Leu
                20

```

## 3. INFORMATION FOR SEQ ID NO:116:

## (1) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 19 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:116:

```

Ala Val Gln Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
1           5           10           15

```

## 4. INFORMATION FOR SEQ ID NO:117:

## (1) SEQUENCE CHARACTERISTICS:

- A. LENGTH:
- B. TYPE:
- C. STRANDEDNESS:
- D. TOPOLOGY:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys  
1                      5                      10                      15

Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO:118:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp Pro Ala Trp Gly Pro  
1                      5                      10                      15

(2) INFORMATION FOR SEQ ID NO:119:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Asp Ile Gly Ser Glu Ser Thr Ile Asp Gln Gln Xaa Ala Val  
1                      5                      10

(2) INFORMATION FOR SEQ ID NO:120:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Ala Glu Gly Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro  
1                      5                      10

(2) INFORMATION FOR SEQ ID NO:121:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

Ser

## (2) INFORMATION FOR SEQ ID NO:122:

## (2) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 15 amino acids  
B) TYPE: amino acid  
C) STRANDEDNESS:  
D) TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO:122:

2 INFORMATION FOR SEQ ID NO:1234

### SEQUENCE CHARACTERISTICS:

- A) LENGTH: 30 amino acids  
B) TYPE: amino acid  
C) STRANDEDNESS:  
D) TOPOLOGY: linear

## X1 SEQUENCE DESCRIPTION: SEQ ID NO: 123

Leu Leu Asn Ser Ser Ala Asp Pro Asn Ala Ser Phe Ala Asn  
20 35 50

## 2 INFORMATION FOR SEQ ID NO. 24:

### SEQUENCE CHARACTERIZATION

- ```

A LENGTH 12 ATOL= 1000
B TYPE 46100 1000
C TRANSDUCERS
D POPULATION 10000

```

## X: SEQUENCE DESCRIPTION: IE: 10 NOV 1994

Asp. Pl. Pr. Asp. Pr. H.L. Gl. Cal. Sep. Pet. Corolla Gls. Style Pistil

## 1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Asp Pro Gly Tyr Thr Pro Gly  
1 5

## 2) INFORMATION FOR SEQ ID NO:126:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Second Residue Can Be Either a Pro or Thr"

x1. SEQUENCE DESCRIPTION: SEQ ID NO:126:

Xaa Xaa Gly Phe Thr Gly Pro Gln Phe Tyr  
1 5 10

## 3) INFORMATION FOR SEQ ID NO:127:

## 1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 9 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS:
- D) TOPOLOGY: linear

## (ix) FEATURE:

D) OTHER INFORMATION: note= "The Third Residue Can Be Either a Gln or Leu"

x1. SEQUENCE DESCRIPTION: SEQ ID NO:127:

Gln Phe Xaa Val Thr Ala Tyr Ala Gln  
1 5

## 4) INFORMATION FOR SEQ ID NO:128:

## 1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 9 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS:

Xaa Xaa Xaa Glu Lys Pro Phe Leu Arg  
 1 5

(2) INFORMATION FOR SEQ ID NO:129:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Xaa Asp Ser Glu Lys Ser Ala Thr Ile Lys Val Thr Asp Ala Ser  
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:130:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Ala Gly Asp Thr Xaa Ile Tyr Ile Val Gly Asn Leu Thr Ala Asp  
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:131:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Ala Ser Ile Ser Gly Ala Glu Lys Ile Thr Thr Thr Ala Ala Glu  
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:132:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Asn Val His Leu Val  
20

## 2. INFORMATION FOR SEQ ID NO:133:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 882 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

## (12) SEQUENCE DESCRIPTION: SEQ ID NO:133:

```

GCAACGCTGT CTTGGGCTTT GCGGTGATCG GTTTCGCGTC GCTGGCGGTG GCGGTGCGCG      60
TCACCATCGG ACCGACCGCG GCTCAAAAC CGCTAGAGCG ACACAAAAAC GCGGAGCGAG      120
GGAACTTCAT GCCCTTCTTG CCGACGCAAC AGCAGGCGCG GGTCCGCGCG GCTCGCGCGG      180
ATGATCGGAC CGCTGGATTC CAGCGCGGGA CCATTCGCGG TGTACAGAAC GTGCTGCTGG      240
GCGCGCGCTAC CTCACCGCGG GTGGGTGGGA GCGCGGCTTC GCTGCGCGCG GAAGCGCGCG      300
CGCTCGCGCG TGTTGTGCTT GCGCGGCTGC CAATCGCGGT CGCGATCATC ATTGCGCGCT      360
TCCCGCGCTT GCAGCGTGGG ATCGCGAGCA TCGCGACCGG ATGCGCGACG AGCGCGGTGA      420
CGACGTGGGG GACGAGCGCG CGACCGAGCG GCGCGACGAC TCGCGTGAGG AGCGCGCGAA      480
CGACCGCGCG GACGAGCGTG CGACCGAGCG TCGCGAGGAC CGCGCGAGCG AGCGCGGTGA      540
CGCGCGCGAG AACGAGCGTG CGCGGAGGCA CGCTCGCGCG GACGAGCGTG GTCGCGAGCA      600
CGCTCGCGCG GACGAGCGTG GTCGAGGCA TCGCGAGGCG GACGAGCGTG GTCGCGAGCG      660
CGCGCGCGCA GCGCGAGGCA CGACCGAGCG AACGAGTGGG AACCGAGGAG GAGCGCGTGG      720
CGCGGAGAGT GTCGCGCGCG GTCGCGAGCG CGCGGAGGCG TCGCGGAGAG CGCGCGCGCG      780
CGCGGAGTTC ATTGCGCGCG GTCGCGAGCG GTCGCGAGCG TCGCGGAGAG CGCGCGCGCG      840
CGCGGAGTTC GCGGTGAGCG GTCGCGAGCG GTCGCGAGCG GTCGCGAGCG GTCGCGAGCG      882

```

## 3. INFORMATION FOR SEQ ID NO:134:

- (1) SEQUENCE CHARACTERISTICS:  
 A LENGTH: 217 base pairs  
 B TYPE: nucleic acid  
 C STRANDEDNESS: single

## X1) SEQUENCE DESCRIPTION: SEQ ID NO:134

CCATCAACCA ACCGCTCGCG CCGCCCGCGC CGCCGGATCC GCGTCCGCG CCACGCCCGC	60
CGGTGCGTCC GGTGCCCCCG TTGCGCGCGT CCGCGCGCTC GCGCGCGACC GGCTGGGTGC	120
CTAGGGCGGT GTTACCGCCC TGGTTGGCGG GGACGCCGCC GGCACCACCG GTACCGCCGA	180
TGCGCGCGTT GCGCGCGCGG GCACCGTTGC CACCGTTGCG ACCGTTGCCA CCGTTCGCCA	240
CCAGCCACCC GCGCGGACCA CCGGCACCGC CGCGCGCGCC CGCACCGCGG GCGTGCCCGT	300
TGCTGCGCGT ACCCGCGGCA CCGCGGTTGC CCGCGTCACC GCGGACGGAA CTACCGCGCG	360
ACCGCGCGTG CCGCGCGCGG TCGCGCGCAC CCGCATTTGC ACCGCGGTCA CCGCGCGCTG	420
GGAGTCCCGC GATTAGGGCA CTGACCGCGG TAACGAGCGC AAGTACTCTC GGTACCGAG	480
CAGTTCCAGA CCACACCACA CCACCGGGTT CTCGGCGGAC TGGGTGAAT GGCAGCCGAT	540
AGCGCGTAGC TGTGCGGTGC GGTCAACCTC GATCATGATC TCGAGGTGAC CGTGACCGCG	600
CGCGCGGAAG GAGGCGGTGA ACTCGCGGTT GAGCGGATCG GCGATCGGTT GGGGCGAGTC	660
CCAGGCCAAT ACCGGGATAC CCGGTGTGTA AGCGCGCGCG AGCGAGGCTT CGGTTGCGCG	720
ACGCTGGTGC GGGTGGGCTG TTACCGCGTT CTCGTCGAAC ACGAGTAGCA GGTCTGCTGC	780
CGCGAGGGCA TCCACCACGC GTTGGCTCAG CTCGT	915

## 2) INFORMATION FOR SEQ ID NO:135

## A) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 1152 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

## 1) MOLECULE TYPE: DNA (genomic)

## A) SEQUENCE DESCRIPTION: SEQ ID NO:135

CTGAGGCTT GCGTACGCTC TCGATCAGCA GATGCTGCGG ATGAGCGCGG TCGGTTGAGC	60
CTGTTGCGAG AAGCACTGCT GAGGATGCTC CCGCGCGAAA CAGGCGCTCA TTGACGCTC	120
TATGAGGCTT TGAAGAGCA GATGCTGCGG GATGCTGATA TCGGACGCTT GCGGTAACAC	180
CTGTTGAGCA TGTGCTGCTT CTGCTGCTGCT GATGCTGCTT CCGGATGAGC TTGTCGCGAG	240
CTGTTGAGCA TGTGCTGCTT CTGCTGCTGCT GATGCTGCTT GATGCTGCTT GATGCTGCTT	300

ACACCCGACG TGTCTACGCG GCGCGGGCTC GGTGAGCAAG TTCACCGCAC CGACGATCCT	480
GCGTTCTGCC TGTCTTAAG CAAGCGGATC GTGTGAGGA AGATCCTGAA TCAGCAGGCC	540
TTGATTGGGG CACACACGTC GGGGCAAGAC GTTGTGAGA GCATCCGCAC GATGAAGCAC	600
TCGCTGGCCT GGGTCGATCG ATCGGGCTCC CTGGCGGAGT TGAACGGGT CGAGGGAAAT	660
GCGCAAAGG CATACTTCAC CGCGCTGGGG CATCTCGTCC CGCAGGAGTT CGCATTCCAG	720
GGCCGCTCGA CTCGGCCGCC GTTGGACGCC TTCAACTCGA TGGTCAGCCT CGGCTATTGG	780
CTGCTGTACA AGAACATCAT AGGGGCGATC GAGCGTCACA GCCTGAACGC GTATATCGGT	840
TTCTTACAC AGGATTCAGG AGGGCAGCGA AGTCTCTGTG CCGAATTCTGG CACGAGCTCC	900
ACTGAAACCG CTGGCCGCTT GTCAGTGGC GTACGTAAT CGGCTGCGCC CAGGCCGCCC	960
CGCCGCGCGA ATACGAGCAG ATCGGACAGC GAATTGCGCC CCAGCCGCTT GGAGCCGCTC	1020
ATACCGGCGG CACTCTACG GCGAGCGAAG AGGCTCGCA CGGTGGCGGC GCGGCTGTC	1080
CGGTCTACTT CGACACGCGC CATCAGCTAG TCACAGCTCG GCGCGACTTC CATTGCCTGC	1140
TTTGGGACG AG	1152

# 2. INFORMATION FOR SEQ ID NO:136:

1. SEQUENCE CHARACTERISTICS:
- A. LENGTH: 655 base pairs
  - B. TYPE: nucleic acid
  - C. STRANDEDNESS: single
  - D. TOPOLOGY: linear

2. MOLECULE TYPE: DNA genomic

## 3. SEQUENCE DESCRIPTION: SEQ ID NO:136:

TTTCTTCCGA TTCTGGAGGG GTTACTTATG ATTCTTTTAN TTGCGATGAG GTTTGAGGAT	1
TGCTATGCGG GAGAGAGCA GTATCTTGCT TACGAGTCTT AGTGGTTTGA TTGAGGAGAT	11
TTCTGTGACT GCGGCGTGGT AGGTCTTTTCT TTCTTCTAGG GTATCTGATG GTTTTCTGGC	21
GTGCTTTTTC GCGCATTAAT GCGGCGGCTT GGTCTGCGCG GCGGATATGG CGGAANGTGG	31
ATGAGGACAG CGGAGATAGG GTTCTGTGCA AGTCTTTTGA GGTCTGCGCG GGGCAGCTTC	41
TTCGCGAATT GTATTAATGA GAATCTTGGT TGTATATGTA TGTGAGCGAA TTGCTGCTGC	51

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

1. SEQUENCE DESCRIPTION: SEQ ID NO: 127:

Asn	Ala	Val	Val	Ala	Phe	Ala	Val	Ile	Gly	Phe	Ala	Ser	Leu	Ala	Val
1				5						10					15
Ala	Val	Ala	Val	Thr	Ile	Arg	Pro	Thr	Ala	Ala	Ser	Lys	Pro	Val	Gln
			20					25					30		
Gly	His	Gln	Asn	Ala	Gln	Pro	Gly	Lys	Phe	Met	Pro	Leu	Leu	Pro	Thr
	35					40						45			
Gln	Gln	Gln	Ala	Pro	Val	Pro	Pro	Pro	Pro	Pro	Asp	Asp	Pro	Thr	Ala
	50					55						60			
Gly	Phe	Gln	Gly	Gly	Thr	Ile	Pro	Ala	Val	Gln	Asn	Val	Val	Pro	Arg
65					70					75					80
Val	Gly	Thr	Ser	Pro	Gly	Val	Ile	Gly	Thr	Pro	Ala	Ser	Pro	Ala	Pro
				85					90					95	
Gln	Ala	Pro	Ala	Val	Pro	Gly	Val	Val	Pro	Ala	Pro	Val	Pro	Ile	Pro
			100					105						110	
Val	Thr	Ile	Ile	Ile	Thr	Pro	Ile	Pro	Gly	Thr	Ala	Pro	Gly	Met	Pro
	115						120						125		
Thr	Ile	Pro	Thr	Ala	Ile	Thr	Thr	Ala	Ala	Val	Thr	Thr	Ser	Ala	Thr
	130					135							140		
Thr	Pro	Pro	Thr	Thr	Ala	Pro	Thr	Thr	Pro	Val	Thr	Thr	Pro	Pro	Thr
145					150					155					160
Thr	Pro	Pro	Thr	Thr	Pro	Val	Thr	Thr	Pro	Pro	Thr	Thr	Pro	Pro	Thr
									165					170	

```

195                200                205
Ala Thr Ala Thr Pro Thr Thr Val Ala Pro Gln Pro Thr Gln Gln Pro
210                215                220
Thr Gln Gln Pro Thr Gln Gln Met Pro Thr Gln Gln Gln Thr Val Ala
225                230                235                240
Pro Gln Thr Val Ala Pro Ala Pro Gln Pro Pro Ser Gly Gly Arg Asn
245                250                255
Gly Ser Gly Gly Gly Asp Leu Phe Gly Gly Phe
260                265

```

ii) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

```

A) LENGTH: 174 amino acids
B) TYPE: amino acid
C) STRANDEDNESS: single
D) TOPOLOGY: linear

```

(ii) MOLECULE TYPE: peptide

xii) SEQUENCE DESCRIPTION: SEQ ID NO:138:

```

Ile Asn Gln Pro Leu Ala Pro Pro Ala Pro Pro Asp Pro Pro Ser Pro
1      5      10      15
Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro
20     25     30
Ser Pro Pro Thr Gly Trp Val Pro Arg Ala Leu Leu Pro Pro Trp Leu
35     40     45
Ala Gly Thr Pro Pro Ala Pro Pro Val Pro Pro Met Ala Pro Leu Pro
50     55     60
Pro Ala Ala Pro Leu Pro Pro Leu Pro Pro Leu Pro Pro Leu Pro Thr
65     70     75     80
Pro Ala Thr Pro Arg Pro Pro Ala Pro Pro Ala Pro Pro Ala Thr Pro
85     90     95
Ala Cys Pro Phe Val Pro Val Pro Pro Ala Pro Pro Leu Pro Pro Ser
100    105    110
Pro Pro Thr Leu Leu Pro Ala Asp Ala Ala Cys Pro Pro Ala Pro Pro
115    120    125
Gly Pro Pro Leu Ala Pro Pro Ser Pro Pro Ala Pro Ser Ala Ala Pro
130    135    140    145

```

Leu Pro Asp Asp Thr Thr Ala Arg Gly Cys Arg Arg Thr Gly  
 165 170

2 INFORMATION FOR SEQ ID NO:139:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 35 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Gln Pro Pro Ala Glu Val Ser Asp Gln Arg Val Ser Gly Leu Thr Gly  
 1 5 10 15  
 Ala Val Gln Pro Ser Pro Arg Thr Thr Ala Glu Asp Pro Arg Pro Arg  
 20 25 30  
 Asn Arg Arg  
 35

2 INFORMATION FOR SEQ ID NO:140:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 104 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Arg Ala Asp Ser Ala Gly Cys Thr Cys Arg Trp Cys Xaa Pro His Glu  
 1 10 15  
 Arg Arg Arg Ala Met Arg Ala Lys His Gly Ser Arg Ser Thr Thr  
 20 25 30  
 Pro Pro Gly Ser Arg His Asp Ser Lys Arg Val Arg Pro Gly Arg Asp  
 35 40 45  
 Pro Pro Trp Ala Gly Ser Ser Asp Val Phe Pro Pro Trp Phe Ala Ala  
 50 55 60  
 Leu Met Pro Ala Arg Arg Val Lys Arg Pro Val Trp Pro Xaa Val Asp  
 65 70 75

Gly Gln Leu Arg Arg Gln Phe Tyr  
100

11 INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(iii) ORIGINAL SOURCE:

- (A) ORGANISM: *Mycobacterium tuberculosis*

(iv) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GGATCCATAT GGGCCATCAT CATCATCATC AGCTGATCGA CATCATCGGG ACC

53

12 INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR Primer"

(iii) ORIGINAL SOURCE:

- (A) ORGANISM: *Mycobacterium tuberculosis*

(iv) SEQUENCE DESCRIPTION: SEQ ID NO:142:

CGTGAATTCG GGGCTGGGTT GGGGCGGGCT CATCTTGAAC GA

42

13 INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR Primer"

(iii) ORIGINAL SOURCE:

GGATCCTGCA GGCTCGAAAC CACCGAGCGG T

31

12 INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

x1 SEQUENCE DESCRIPTION: SEQ ID NO:144:

GGTGAATTC AGCGCTGGAA ATCGTCGCGA T

32

13 INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

x1 SEQUENCE DESCRIPTION: SEQ ID NO:145:

AGATCGAACC CTCAGATGAA CACCGATGCC GCT

33

14 INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "PCR primer"

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mycobacterium tuberculosis

## (2) INFORMATION FOR SEQ ID NO:147:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1993 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

AN ORGANISM: Mycobacterium tuberculosis

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 152..1273

## (x) SEQUENCE DESCRIPTION: SEQ ID NO:147:

```

TGTTCCTTCGA CGGCAGGCTG GTGGAGGAAG GGCCACCGA ACAGCTOTTC TCGTGGCCGA      60
AGCATGCGGA AACGCGCCGA TACGTGCGCG GACTGTGCGG GGACGTCAAG GACGCCAAGC      120
CCGGAATTG AAGAGCACAG AAAGGTATGG C GTG AAA ATT CGT TTG CAT ACG      180
                               Val Lys Ile Arg Leu His Thr
                               1           5

GTG TTG GCG GTG TTG ACC GCT GCG GCG GTG GTG CTA GCA GCG GCG GCG      240
Leu Leu Ala Val Leu Thr Ala Ala Pro Leu Leu Leu Ala Ala Ala Gly
          10           15           20

GGT GCG TCG AAA CGA GCG ACC GGT TCG GCT GAA ACC GCG GCG GCG GCG      300
Gly Gly Ser Lys Pro Pro Ser Gly Ser Pro Gly Thr Gly Ala Gly Ala
          25           30           35

GGT ACT GTG GCG ACT ACC GCG GCG TCG TCG GCG GTG ACC TTG GCG GAG      360
Gly Thr Val Ala Thr Thr Pro Ala Ser Ser Pro Val Thr Leu Ala Gly
          40           45           50           55

AGG GGT AGG AGG TTG TTG CAG GCG GTG TTG AAA TTG TCG GGT GCG GGT      420
Arg Arg Ser Thr Leu Leu Thr Tr Leu Thr Arg Leu Thr Thr Pro Ala
          60           65           70           75

TTT CAG GAG AGG TAT GCG AAC GTG ACG ATT ACC GGT CAG GCG ACC GGT      480
Phe His Gly Arg Tyr Pro Asn Val Thr Ile Thr Ala Gly Gly Thr Gly
          80           85           90           95

TTT GGT GTG GCG ATT GCG CAG GCG GTG GTT GCG ACC GTG AAC ATT GCG      540
Phe Gly Ala Gly Trp Asn Gly Ala Ala Ala Gly Thr Val Asn Ile Gly
          100           105           110           115

```

CTG ATG AAC ATC GCG CTA GCC ATC TCC GCT CAG CAG GTC AAC TAC AAC Leu Met Asn Ile Ala Leu Ala Ile Ser Ala Gln Gln Val Asn Tyr Asn 120 125 130 135	556
CTG CCC GGA GTG AGC GAG CAC CTC AAG CTG AAC GGA AAA GTC CTG GCG Leu Pro Gly Val Ser Glu His Leu Lys Leu Asn Gly Lys Val Leu Ala 140 145 150	604
GCC ATG TAC CAG GGC ACC ATC AAA ACC TGG GAC GAC CCG CAG ATC GCT Ala Met Tyr Gln Gly Thr Ile Lys Thr Trp Asp Asp Pro Gln Ile Ala 155 160 165	652
GCG CTC AAC CCC GGC GTG AAC CTG CCC GGC ACC GCG GTA GTT CCC CTG Ala Leu Asn Pro Gly Val Asn Leu Pro Gly Thr Ala Val Val Pro Leu 170 175 180	700
CAC CCC TCC GAC GGG TCC GGT GAC ACC TTC TTG TTC ACC CAG TAC CTC His Arg Ser Asp Gly Ser Gly Asp Thr Phe Leu Phe Thr Gln Tyr Leu 185 190 195	748
TCC AAG CAA GAT CCC GAG GCG TGG GGC AAG TCG CCC GGC TTC GCG ACC Ser Lys Gln Asp Pro Glu Gly Trp Gly Lys Ser Pro Gly Phe Gly Thr 200 205 210 215	796
ACC CTC GAC TTC CCC GCG CTC CCC GGT GCG CTC GGT GAG AAC GGC AAC Thr Val Asp Phe Pro Ala Val Pro Gly Ala Leu Gly Glu Asn Gly Asn 220 225 230	844
GGC CCC ATG CTG ACC GGT TCG GCG GAG ACA CCC GCG TCG GTG GCG TAT Gly Gly Met Val Thr Gly Lys Ala Glu Thr Pro Gly Lys Val Ala Tyr 235 240 245	892
ATG CCC ATC AGC TTC CTC GAC CAG CCC AGT CAA CCC GGA CTC GCG GAG Ile Gly Ile Ser Phe Leu Asp Gln Ala Ser Gln Arg Gly Leu Gly Gln 250 255 260	940
CCC CAA CTA GCG AAT ACC TCT GCG AAT TTC TTC TTC CCC GAC GCG CAA Val Gln Leu Gln Asn Ser Ser Gly Asn Pro Leu Leu Pro Asp Ala Gln 265 270 275	988
ATG ATT CAG GCG TTC GCG GGT GGT CTC GAG AAA ACC CTC GCG AAT Ser Ile Gln Ala Ala Ala Ala Gln Phe Ala Ser Lys Thr Pro Ala Asn 280 285 290 295	1036
CAG GCG ATT TCG ATG ATG GAC GCG GCG GCG GCG GAG GCG TAC CCC ATC His Ala Ile Ser Met Ile Asp Gly Pro Ala Pro Asp Gly Tyr Pro Ile 300 305 310	1084
ATG AAT TAC CAG CAG GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT Met Asn Tyr CAG CAG GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400	1132

INFORMATION FOR SEQ ID NO. 148

## 2. SEQUENCE CHARACTERISTICS

- A LENGTH: 374 amino acids  
B TYPE: amino acid  
C TOPOLOGY: alpha

... ..

NO.	FREQUENCY	DESCRIPTION	AGE	SEX	RACE
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					
81					
82					
83					
84					
85					
86					
87					
88					
89					
90					
91					
92					
93					
94					
95					
96					
97					
98					
99					
100					

Al Asp Ile Arg Ser His Thr Leu Leu Ala Val Ser Thr Ala Ala Pro  
1 5 10

Leu Ser Leu Ala Ala Ala Gly Asp Gly Ser Leu Pro Pro Ser Ile Thr  
15 20

50                                      55                                      60  
 Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn Val Thr  
 65                                      70                                      75                                      80  
 Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala  
 85                                      90                                      95  
 Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly  
 100                                      105                                      110  
 Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser  
 115                                      120                                      125  
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys  
 130                                      135                                      140  
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr  
 145                                      150                                      155                                      160  
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro  
 165                                      170                                      175  
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr  
 180                                      185                                      190  
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly  
 195                                      200                                      205  
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly  
 210                                      215                                      220  
 Ala Leu Gln Gln Asn Gly Asn Ile Ile Met Val Thr Gly Cys Ala Gln  
 225                                      230                                      235                                      240  
 Thr Pro Gly Lys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala  
 245                                      250                                      255  
 Ser Gln Arg Gly Leu Gly Gln Ala Gln Leu Ile Asn Ser Ser Gly Asn  
 260                                      265                                      270  
 Thr Ser Leu Thr Asn Ala Gln Ser Thr Gln Ala Ala Ala Ser Thr  
 275                                      280                                      285  
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro  
 290                                      295                                      300  
 Ala Phe Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn  
 305                                      310                                      315                                      320  
 Arg Arg Ser Thr Asp Ala Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr

His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu  
 355 360 365

Ile Ala Thr Ile Ser Ser  
 370

(2) INFORMATION FOR SEQ ID NO:149:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1993 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:149:

TGTTTCTTCGA CGGCAGGCTG GTGGAGGAAG GGGCCACCGA ACAGCTGTTC TGCTCCCGA	50
AGCATGCGGA AACCSCCCGA TACSTGCGCG GACTGTGCGG GGACGTCAAG GACGCCAAGC	120
CGCGAAATTC AAGAGCACAG AAAGGTATGG CTTGAAAAAT CGTTTCCATA CGGTGTGGG	180
CGTGTGACG GCTGCGCGCG TGCTCTTAGC AGCGGCGGGC TGTGGCTCGA AACCACCGAG	240
CGTTTCCGCT GAAACGGGCG CGGCTCCCGG TACTGTGCGG ACTACCGCGG CGTGTGCGG	300
GTGACGTTG GCGGAGACCG GTAGTACGCT GCTGTACCGG CTGTTCAAGC TGTGGGGTCC	360
CGCGTTTCAG CAGACGTATC CGAAGCTCAC GATCACCGCT CAGGGCACCG GTTCTGGTGC	420
CGCGATCGCG CAGCGCGCGG CGCGACGCT CAAGTTTGGG CGCTCCGAGC GTATCTGTG	480
CGAAGCTGAT ATCGCGCGCG AGAAGCGGCT GATCAACATC CGCTAGCGA TGTCCGCTCA	540
CGCGGTCAAG TACAACCTGC CGCGAGTCAG CGAGCACCTG AAGCTGAAGC GAAAAGTCCT	600
CGCGGCGATG TACCGGGCGA CGATCAAAAT CTGGGACGAC CGCGAGATCG CTGCGCTCAA	660
CGCGGCTGCT GAGCTGCGG CGAGGCGGCT ATTTGCTGTC TACGCTCGG AAGGCTCGG	720
CGAGGCTGCT CTGTTGAGCT AGTACGCTG CAAGGAGATC CGAGGCGGCT CGCGAGCTG	780
CGCGGCTGCT CGAGGAGCTC CGAGGCTGCT CGAGGCTGCT CGAGGCTGCT CGAGGCTGCT	840
CGAGGCTGCT ATTTGAGCTG GTTGGGCGGA TACCGCGCGG TGCTGCGCT ATATCGGCAT	900
CGAGGCTGCT GAGGAGCTG CTCAAGCGG AGCTCGCGCA CGCGAATAG CGAATAGCTG	960
CGAGGCTGCT CTGTTGAGCT AAGGAGGAG CAAGGAGGAG CGCGGCTGCT CGTTGAGCTG	1020

GACGTTTCAG GCATTTCTGC ACTGGGCGAT CACCGACGGC AACAGGGCTT GATTCTCGA 1200  
 CCAGGTTTCAT TTCCAGCCGC TCCCCCCCGC GGTGGTGAAG TTGTCTGACG CGTTGATCGC 1260  
 GACGATTTCC AGCTAGCCTC GTTGACCACC ACGCGACAGC AACCTCCGTC GGGCCATCGG 1320  
 GCTGCTTTGC GGAGCATGCT GCGCCGTGCC GGTGAAGTCG GCCGCGCTGG CCGGGCCATC 1380  
 CGGTGGTTGG GTGGGATAGG TGCGGTGATC CCGCTGCTTG CGCTGGTCTT GGTGCTGGTG 1440  
 GTGCTGGTCA TCGAGGCGAT GGGTCCGATC AGGCTCAACG GGTGCAATT CTTCACCGCC 1500  
 ACCGAATGGA ATCCAGGCAA CACCTACGGC GAAACCGTTG TCACCGACCC GTGGCCGATC 1560  
 CGGTCCGGCC CTACTACGGG GCCTTCCCGC TGATCCTCGG GACGCTGGCG ACCTCGGCAA 1620  
 TCCCCCTCAT CTTCCCGGTC CGGCTCTCTG TACGAGCCGC GTGGGTGATC GTGGAACGGC 1680  
 TCCCGAAACG GTTGGCCGAG GCTGTGGGAA TAGTCTCGGA ATTGCTCGCC GGAATCCCCA 1740  
 GCCTGCTGCT CGGTTTGTGG GGGGCAATGA CGTTCCGGCC GTTCATCGCT CATCACATCG 1800  
 CTCGGTGTAT CGCTCACAAC GCTCCCGATG TCCCGGTGCT GAACTACTTG CGCGGCGACC 1860  
 CGGCAACCGG CGAGGGCATG TTGCTCTCCG GTCTGTGTTT GCGCGTGATG GTCTTTCCCA 1920  
 TTATCGCCAC CACCACTCAT GACCTSTTCC GGCAGGTGCC GGTGTTGCCC CCGGAGGGCG 1980  
 GATCGGGAA TTC 1993

# II. INFORMATION FOR SEQ ID NO.150:

## A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 374 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

## xi. SEQUENCE DESCRIPTION: SEQ ID NO. 15

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro  
 1 10  
 Leu Leu Leu Ala Ala Ala Gly Tyr Gly Ser Lys Lys Thr Ser Gly Ser  
 20 30  
 Phe His Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser  
 35 45  
 Met Phe Thr Thr Thr Ala His Thr Gly Ser Thr Leu Leu Tyr Pro Leu

Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala  
 85 90 95  
 Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly  
 100 105 110  
 Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser  
 115 120 125  
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys  
 130 135 140  
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr  
 145 150 155 160  
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro  
 165 170 175  
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr  
 180 185 190  
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly  
 195 200 205  
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly  
 210 215 220  
 Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu  
 225 230 235 240  
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala  
 245 250 255  
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn  
 260 265 270  
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe  
 275 280 285  
 Ala Ser Lys Thr Thr Ala Asn Gln Ala Ile Ser Met Ile Asp Lys Thr  
 290 295 300  
 Ala Pro Arg Gly Tyr Val Ile Gln Asn Tyr Gln Tyr Ala Ile Val Asn  
 305 310 315 320  
 Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu  
 325 330 335  
 His Thr Ala Ile Thr Asp Gly Asn Lys Ala Ser Pro Leu Asn Gln Val  
 340 345 350

370

## (2) INFORMATION FOR SEQ ID NO:151:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1777 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ix) SEQUENCE DESCRIPTION: SEQ ID NO:151:

```

GGTCTTGACC ACCACCTGGG TGTGGAAGTC GGTGCCCCGA TTGAAGTCCA GGTACTCGTG      60
GGTGGGGCGG GCGAAACAAT AGCGACAAGC ATGCGAGCAG CCGCGGTAGC CTTTGACGGT      120
GTAGCGAAAC GGCAACGCGG CCGCGTTGGG CACCTTCTTC AGCGCTGATT TGCACAACAC      180
CTCGTGGGAG GTGATGCGGT CGAATTGTGG CCGCGCAACG CTGCGGACCA GCGCGATCGG      240
CTGCAACCGG GCAGCGCGCG TCGTCAACCG GCATCCCGTT CACCGCGACG GCTTGCGCGG      300
CGCAACGCAT ACCATTATTC GAACAACCGT TCTATACTTT GTCAACGCTG GCGCGTACCG      360
AGCGCGCGAC AGGATGTGAT ATGCCATCTC TGCGCGCACA GACAGGAGCC AGGCCTTATG      420
ACAGCATTCG GCGTCGAGCC CTACGGGCAG CGGAAGTACC TAGAAATCGC CCGGAAGTGC      480
ATCGCGTATA TCGACGAAGC CAAGGGTGAC GCGATCGTCT TTCAGCACGG CAACCGCCAC      540
TCTCTTTACT TGTGGCGCAA CATCTMTGCG CACTTGBAAG GGCTGGGCGG GCTGGTGGCG      600
TGGGATCTCA TCGGATGGG CGCTCGCGAC AGCTCAGCG CATCGGGACC CGACCGTAT      660
AGCTATCGCG AGCAACGAGA CTCTTTCTTC GCGCTGTGGG ATCGCGTCCA CTTCGGCGAC      720
CAGCTGGTAC TGGTGTGSCA CGAATGGGGG TCGGCGCTCG GCTTCGACTG GCTTAACGAG      780
CATCGCGAAC GATGCGAGGG GATCGCGTTC ATGGAAGCGA TCGTCACCGG GATGACGTGG      840
TGGGATTCGG TCGCGGCTGT TCGGCTGTTC TTGGAAGCTT TCGGATTCGG TGAAGCGGAG      900
TGAATTCGCT TCGAGCAGAA CATCTTCTTC GAAGCTGTTC TCGGCGCGCG GATCTTCTCA      960
GAGCTGAGCG AGAAGGAAT GAAGCACTAT CGCGCGGAG TCGTGAGCGG TCGGAGTAC      1020
CTTCGCGTCA GCTTGTCTTC TCGAGGAAT CTTCGAATCG AGCTGAGCGG TCGCGAGGTC      1080
CTCGCTTTCG TCGAGGAGTA TCGGAGGTGG CTGAGGAGAA TCGGATGCTT GAAGCTCTTC      1140
ATGAAGCGCG AGCTGCGCGG GATGATCACT GCGCGGCTTC CTGATCTATCT GAGGAGCTGG      1200

```

GACCAAGAAT GTGATTTCCG GCGAAGGCGG CGCCCTGCTT GTCAACTCAT AAGACTTCCT 1380  
 GCTCCGGGCA GAGATTCTCA GGGAAAAGGG CACCAATCGC AGCCGCTTCC TTGCAACGA 1440  
 GGTGACAAA TATACGTGGC AGGACAAAGG TCTTCCTATT TGCCAGCGA ATTAGTCGCT 1500  
 GCCTTTCTAT GGGCTCAGT CGAGGAAGCC GAGCGGATCA CGCGTATCCG ATTGGACCTA 1560  
 TGAACCGGT ATCATGAAAG CTTGGAATCA TTGGAACAGC GGGGGCTCCT GCGCCGTCCG 1620  
 ATCATCCCAC AGGGCTGCTC TCACAACGCC CACATGTACT ACCTTTACT AGCGCCAGC 1680  
 GCCGATCGGG AGGAGGTGCT GCGCGCTCTG ACGAGCGAAG GTATAGGCCG GGTCTTTTAT 1740  
 TACGTGCCGC TTCACGATTC GCGGGCCGGG CGTCGCT 1777

2. INFORMATION FOR SEQ ID NO:152:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 324 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GAGATTGAAT CGTACCGGTC TCGTTAGCGG CTGCGTCCCG TGAATGCCCA TATCAGGCAC 50  
 GCGATGTTTC TGGGTGTCCA CTTTCGCCCC ATGCCCGGAC GTTGGTAAAC CCAGGGTTTG 100  
 ATGATTAATT CGCGGGGACC TTTCGGGGA AGCGGCCAGG ATGTGGGTGA GCGCGGGGCT 150  
 TTTTGTGGG CAJGGBACCG CTGATGCTC AGCCCCGCTG GCGCGAGCTA GCGAGGTTT 200  
 TTTTGTGTC GTGACAGTC GTACTCGGCT GACGACCGCG GCGGGTGCTT GGTGAAGAC 250  
 GGTGACGAC GCGGCGGATT CAGA 324

INFORMATION FOR SEQ ID NO:153

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 324 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:153

CAATAAGG CGCGTTGCTT TCGAAGGGA CTTGTAGCA GTCAAGCACT TCGCGTCCCG 50

```

GCACGACTTC CAGCCCGACT CGATCGGCGT GGTGACCCGT CCTGTCCCTA TGGCTGCCGTG      240
GGTAGCTCCG GTTCGGAAGC GATTTCGGTT CCTCACTGAC CTCGACGCCG ACGAGCAGCG      300
GTGGGCCCGC TCGACGAAC GGCACCGCCG CGAAGTGGAG AACCGCTGG CGGTGCTGCG      360
GTCTGATCA ACCTGCCGGC GATCGTGCCG TTCCGCTGGC ACGGTTGCGG CTGGACGCGG      420
CTGAATCGAC TAGATGAGAG CAGTTGGGCA CGAATCCGGC TGTGCTGGTG AGCAAGACAC      480
GAGTACTGTC ATCACTATTG GATGCACTGG ATGACCGCCG TGATTCAGCA GGACCAATGG      540
AATGCCCCGG GGCAAAACGT CTCGGAGATG ATCGGCGTCC CCTCGGAACC CTGCGGTGCT      600
GGCTTCATTG GGACATCGGT CCGGCTCGCG CGATCGTGGT GACGCCAGCG CTGAAGGAGT      660
GGAGCGCGGC GGTGGACCGC TTCTGGAC CCCTCCAGAC GGTGCTGCTG CTTAAGCCCG      720
GGATCGCGCA GAAGCGCTTC GAGGTGCGCG CGAAGAGATT TTGTTGTTT CCGACCGTCC      780
CGCACAGCCA CGCCGAGCGG GTTCGCGCCG AGCACCAGCA CTTGCTGGGG CCGGCGCGCG      840
CGCACAGCAC CGACGAGTGT GTGCTACTGG GCGCCGAGC GAAAGTTGTT GCGGCACTGC      900
CGGTTAACC GCGAGAGGGT CTGACGCGCA TCGAGGATGT GCACATCTGG ACCGCGAGT      960
CGGTGCGCGC CGACCGGCTC GACTTTGCGC CGAAGCACAA ACTGGCCGTC TTGGTGGTGT      1020
CGGCGATCCC GGTGCGCGAG CCGGTGCGCG TGGCGGCTAG GCGCGAGTAG GCGGCTTCCA      1080
CGAGCTGGGT CGAGCTGCGC GTGACGCGCA CGTTGCGCGC GCGGCTGCA GCGAGGCGCG      1140
CGGTGCGCGA GGTGCGCGCG CGGTGCGCGC AGCGGCTGCG TTGAGTGGCG CGCATGCTTT      1200
CGGTGCGAGC TGTAGCGCGA GTGCGCGCTG CGAATGATGT GCTGTGCGTT CGGTGCGTGC      1260
CGGCTGCAAT TGAAGCGCGC CGCACAGCA CGATTGCGCG CGCATGCTTC CCGGCGCGCG      1320
CGGCTGCGCG CTACAGCG

```

SEQUENCE INFORMATION FOR SEQ. NO. 154

SEQUENCE CHARACTERISTICS

- A LENGTH 131 base pairs
- B TYPE nucleic acid
- C STRANDEDNESS single
- D TOPOLOGY linear

SEQUENCE DESCRIPTION FOR SEQ. NO. 154

SEQUENCE: GCGGCGAGCG GCGGCGAGCG GCGGCGAGCG GCGGCGAGCG GCGGCGAGCG GCGGCGAGCG

```

GCACCGGCGG TGGCGGGGGC GCCGGCAACG ACCCGGGCAG CACCGGCAAT CCGGGCGGTA      240
AGGGCGGGCA CGGCGGGATC GCGGGTGCCG CCGGGGGCGG GGGCGCGGCC GGCACCGGCA      300
ACGGCGGGCA TGGCGGCAAC C                                          321

```

## (2) INFORMATION FOR SEQ ID NO:155:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 492 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## X1) SEQUENCE DESCRIPTION: SEQ ID NO:155:

```

GAAGACCCCG CCCCCCATA TGGATCGGCT GGGCGACTAG TTTCGGCGAA GGTGCACGCG      50
CGGGCGTCCG GCTGATGATC ACCGGTGGCT ACCGGCGCAA CCGCACCGGA TGGGTGCTGC      100
CGTTGGCTTC CGAACTGCTC AATTGGGCGG AAGCGGAGCG GCACCGCGGA ATGACGAGCG      150
CGGTCCACGA TTGGGTGCA AAGATGCTGC TGCAAATGCT GCACCGCGGA GGTACGCT      200
ATCACCGCACT TCGGTCAGC GCTCGCGGA TCAAGGCGCG GATCACCGCG TTTCGTCCGC      300
TACCACTATC GGCTCGCGGG CTCGAAGCGA GCATCGCGGA TTTCGGCGCG TGGGTCAGT      350
TCCCCCGCGA TCGCGGCTAC GACGGCTCG AAATCATGCG CAGCGAAGGG TATCTGCTCA      400
ATCACTGCTT GCGCGCGCGG ACCAACAAGC GCACCGAATC GTGGGGCGCG ACACCGGCGA      450
ATCTTGGCGG GT                                          492

```

## 2) INFORMATION FOR SEQ ID NO:156

## 1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 556 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS
- D) TOPOLOGY: linear

## X1) SEQUENCE DESCRIPTION: SEQ ID NO:156

```

Pro Ala Thr His Leu Val Met Ile Asp Ala Val Ser Leu Pro Arg Ala      1
5
Asp Ala Ala Asp Val Ala Asp Pro Leu Thr Pro Thr Ser Ala Arg Arg      15

```

50	55	60
Ser Ala Ala Gln Asp Val Ile Cys His Leu Cys Pro His Arg Gln Glu		
65	70	75 80
Pro Gly Leu Met Thr Ala Phe Gly Val Glu Pro Tyr Gly Gln Pro Lys		
	85	90 95
Tyr Leu Glu Ile Ala Gly Lys Arg Met Ala Tyr Ile Asp Glu Gly Lys		
	100	105 110
Gly Asp Ala Ile Val Phe Gln His Gly Asn Pro Thr Ser Ser Tyr Leu		
	115	120 125
Trp Arg Asn Ile Met Pro His Leu Glu Gly Leu Gly Arg Leu Val Ala		
	130	135 140
Cys Asp Leu Ile Gly Met Gly Ala Ser Asp Lys Leu Ser Pro Ser Gly		
	145	150 155 160
Pro Asp Arg Tyr Ser Tyr Gly Glu Gln Arg Asp Phe Leu Phe Ala Leu		
	165	170 175
Trp Asp Ala Leu Asp Leu Gly Asp His Val Val Leu Val Leu His Asp		
	180	185 190
Trp Gly Ser Ala Leu Gly Phe Asp Trp Ala Asn Gln His Arg Asp Arg		
	195	200 205
Val Gln Gly Ile Ala Phe Met Glu Ala Ile Val Thr Pro Met Thr Trp		
	210	215 220
Ala Asp Trp Pro Pro Ala Val Arg His Val Phe Gln Gly Phe Arg Ser		
	225	230 235 240
Pro Gln Gly Glu Pro Met Ala Leu Glu His Asn Ile Phe Val His Arg		
	245	250 255
Val Leu Pro Gly Ala Ile Leu Arg Gln Leu Ser Asp His Gln Met Asn		
	260	265 270
His Tyr Arg Arg Arg Val Val Asn Gln His Gln Asp Arg Ala Val Thr		
	275	280 285 290
Leu Ser Thr Pro Asp Asn Leu Pro Ile Asn Gly Glu Pro Ala His Val		
	295	300 305
Val Ala Leu Val Asn Gln Tyr Arg Ser Trp Leu Glu Glu Thr Asp Met		
	310	315 320 325
Trp Leu Leu Phe Ile Arg Ala Gln Pro Thr Thr Thr Thr Thr Thr		

Pro Gly Val His Phe Val Gln Glu Asp Ser Asp Gly Val Val Ser Trp  
 355 360 365  
 Ala Gly Ala Arg Gln His Arg Arg Pro Gly Ser Ala Leu Ile Ser Arg  
 370 375 380  
 Asp Gln Glu Cys Asp Phe Arg Arg Arg Arg Arg Pro Ala Cys Gln Leu  
 385 390 395 400  
 Ile Arg Leu Pro Ala Pro Gly Arg Asp Ser Gln Gly Lys Gly His Gln  
 405 410 415  
 Ser Gln Pro Leu Pro Ser Gln Arg Gly Arg Gln Ile Tyr Val Ala Gly  
 420 425 430  
 Glu Arg Ser Ser Tyr Leu Pro Ser Glu Leu Val Ala Ala Phe Leu Trp  
 435 440 445  
 Ala Gln Phe Glu Glu Ala Glu Arg Ile Thr Arg Ile Arg Leu Asp Leu  
 450 455 460  
 Trp Asn Arg Tyr His Glu Ser Phe Glu Ser Leu Glu Gln Arg Gly Leu  
 465 470 475 480  
 Leu Arg Arg Pro Ile Ile Pro Gln Gly Cys Ser His Asn Ala His Met  
 485 490 495  
 Tyr Tyr Val Leu Leu Ala Pro Ser Ala Asp Arg Glu Glu Val Leu Ala  
 500 505 510  
 Arg Leu Thr Ser Glu Gly Ile Gly Ala Val Phe His Tyr Val Pro Leu  
 515 520 525  
 His Asp Ser Pro Ala Gly Arg Arg  
 530 535

# C INFORMATION FOR SEQ ID NO:157

## SEQUENCE CHARACTERISTICS

- A LENGTH: 584 amino acids
- B TYPE: amino acid
- C STRANDEDNESS:
- D TOPOLOGY: linear

## XI SEQUENCE DESCRIPTION SEQ ID NO:157

Asp Glu Ser Ala Pro Arg Ser Pro Met Leu Phe Ser Ala Arg Pro Arg  
 1 5

...  
 530 535

Gly Asp Asp Arg Ala Gly Leu Gly Val Asp Glu Gln Phe Arg His Val  
 50 55 60  
 Gly Phe Leu Glu Pro Ala Pro Val Leu Val Asp Gln Arg Asp Asp Leu  
 65 70 75 80  
 Gly Gly Leu Thr Val Asp Trp Lys Val Ser Trp Pro Arg Gln Arg Gly  
 85 90 95  
 Ala Thr Val Leu Ala Ala Val His Glu Trp Pro Pro Ile Val Val His  
 100 105 110  
 Phe Leu Val Ala Glu Leu Ser Gln Asp Arg Pro Gly Gln His Pro Phe  
 115 120 125  
 Asp Lys Asp Val Val Leu Gln Arg His Trp Leu Ala Leu Arg Arg Ser  
 130 135 140  
 Glu Thr Leu Glu His Thr Pro His Gly Arg Arg Pro Val Arg Pro Arg  
 145 150 155 160  
 His Arg Gly Asp Asp Arg Phe His Glu Arg Asp Pro Leu His Ser Val  
 165 170 175  
 Ala Met Leu Val Ser Pro Val Glu Ala Glu Arg Arg Ala Pro Val Val  
 180 185 190  
 Gln His Gln Tyr His Val Val Ala Glu Val Glu Arg Ile Pro Glu Arg  
 195 200 205  
 Gln Gln Lys Val Ser Leu Leu Ala Ile Ala Ile Ala Val Gly Ser Arg  
 210 215 220  
 Trp Ala Glu Leu Val Arg Arg Ala His Pro Asp Gln Ile Ala Gly His  
 225 230 235 240  
 Gln Pro Ala Gln Pro Phe Gln Val Arg His Asp Val Ala Pro Gln Val  
 245 250 255  
 Arg Arg Arg Gly Val Ala Val Leu Val Ser Arg Ala Val Thr Leu Ala  
 260 265 270  
 Arg Val Asp Ile Arg His Leu Leu Arg Gly Asp Phe  
 275 280

1 INFORMATION FOR SEQ ID NO 158

1 SEQUENCE CHARACTERISTICS  
 A LENGTH 280 base pairs  
 B TYPE DNA

ATGAACATGT CGTCGGTGGT GGGTCGCAAG GCSTTTGCGC GATTCGCGCG JTACTGCTCG	60
GGCATGCACG CGATCGCGCG TTTCTCCGAT GCSTTGCSCC AAGAGCTGCC GGGTAGCGGA	120
ATCGCGCTGT CGGTGATCCA CCGGCGCTG ACCGAGACAC CGCTGTTGGC CAACGTCGAC	180
CGCGCGGACA TGCGCGCGCG GTTTCGCAGC CTCAGGCGCA TTCCCGTTCA CTGGGTGCGG	240
GCAGCGGTGC TTGACGGTGT GGCG	264

2. INFORMATION FOR SEQ ID NO:159:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:159:

TACTGCGCA CGATGACGTC GCGGTCCAGG CCGACCGCTT CAAGCACCAG CCGGACCAGC	60
AAGCGCGTGC GATCGTTAGC CCGGAAGCAG TGGGTGAGCA CCGGGCGTCC GCGCGCAAC	120
AGTGTGAGCA CACGATGTAG CGCGCGCTGT GCTCCATTCC GCGTTGGGAA TTGGCGATAC	180
TTCTGCTTCA TGTAGCGCGT GCGCGCTCA TTTATCGACT GGCTGGATTG CCGCGACTCG	240
CGCTTGAGCA GGTGATGGT TAGCAGGCTC TTGAATGCGG TTTCGTGCGG CGCTGAGTGC	300
TTCTGCTTAT ATCGGCTGAG TTGGCGAAG CGGAGGAGCT GAAGCTGAT CCGGTCCGCA	360
AAGGTGCTG GAGCGCGCG CCGAAGCTCC CGGATGAGC CGAGGTGCGT AAGTGGCTG	420
ATCGGTAGCC CCGCGAGCGT TCGCTCTGCT CGGAATTTC CGAGGAGGCT CGGTAGCCAC	480
CGGCTATGAC TAAGCAACCG TTGCGAGTA CGGATGCTCA TTTCGCGATC CCGCAGACCA	540
CGGCTATGAC TAAGCAACCG TTGCGAGTA CGGATGCTCA TTTCGCGATC CCGCAGACCA	600
CGGCTATGAC TAAGCAACCG TTGCGAGTA CGGATGCTCA TTTCGCGATC CCGCAGACCA	660
CGGCTATGAC TAAGCAACCG TTGCGAGTA CGGATGCTCA TTTCGCGATC CCGCAGACCA	720
CGGCTATGAC TAAGCAACCG TTGCGAGTA CGGATGCTCA TTTCGCGATC CCGCAGACCA	780
CGGCTATGAC TAAGCAACCG TTGCGAGTA CGGATGCTCA TTTCGCGATC CCGCAGACCA	840
CGGCTATGAC TAAGCAACCG TTGCGAGTA CGGATGCTCA TTTCGCGATC CCGCAGACCA	900

CGCACCACCG TCGGTTCCGA CGTACGGACC GGGTCCGACA CCATGTTTCT GGCCCCAGTA 1080  
 ACCATCGGCG ACGGCGCGTA TACCGGGGCC GGCACAGTGG TCGGGGAGGA TGTCCCGCCG 1140  
 GGGGCGCTGG CAGTGTCCGC GGGTCCGCAA C 1171

(2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 227 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:160:

JCAAAGGCGG CACCGGCGGG GCGGSCATGA ACAGCCTCGA CCGGCTGCTA GCGCCCCAAG 60  
 AGGCGGCGCA AGGCGGCGAC GCGGSCACCG GCGGCAACCG GCGGCGCGCG GGCACCAAGT 120  
 TCACCAAGG GCGGACGGG AACGCGGCGA ACGGCGGTGA GCGGCGGCTC GCGGSCAACG 180  
 GCGGAAACGG GCGGAAACGG GCAGACAACA CCACCAACCG GCGGCGCG 227

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 304 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:161:

CTGCGGACG ATGCGGCGGG AGGCGGCTAG CGTGGGCGCG GGTGTACCT CAGGCGCGAA 60  
 GCGGCGCGAC GCGTTCAGTC CAACCAAGCG GCGGACCGCG GCGGACCGCG GCAACCGCGG 120  
 TACTGCGAA CTGCTCGGCG GCAACCGCGG TCACGCGCGG AATGCGGCGA ACGGCGCGAG 180  
 TCTGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG 240  
 TCTGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG 304

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 base pairs
- (B) TYPE: nucleic acid



(D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:163:

GGGCGGCGG GGGCGGATTT TCTCGTCCT TGATTGTCG TGGGGATAAC GCGCGTGATG	60
GTGGTAACGG CCGGATGGG GGGGCTGGG GGGCTGGCG CCGCGGCGG GCGCGCGGCC	120
TGATCAGCCT GCTGGGCGG CAAGGCGCG GCGGGCGCG CCGGACCGG GGGCGGCGG	180
GTGTTGGCGG TGACGGCGG GCGGGCGG CCGGCAACCA GGCCTTCAAC GCAGGTGCCG	240
GCGGGCGCG GCGGCTGAT AGCCTGCTG GCGGCCAAG CCGCGGCGG GCGGGCGGA	300
GCGGGCGG GCGCGGTSTT GCGGTGAC	329

2) INFORMATION FOR SEQ ID NO:164:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:164:

GCAACGGTGG CAACGGCGG ACCAGCAGCA CGGTGGGGAT GCGCGGAGGT AACTGTGGTG	60
GCGCGGGGCT GATCGGCAAC	80

2) INFORMATION FOR SEQ ID NO:165:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 192 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:165:

AGTCTTCTG TACTGAGG GAGGAGTC GGTACGTTG GCGGCGAAT ATGAGCTTA	60
AGTCTTCTG TACTGAGG GAGGAGTC GGTACGTTG GCGGCGAAT ATGAGCTTA	120
AGTCTTCTG TACTGAGG GAGGAGTC GGTACGTTG GCGGCGAAT ATGAGCTTA	180
AGTCTTCTG TACTGAGG GAGGAGTC GGTACGTTG GCGGCGAAT ATGAGCTTA	240
AGTCTTCTG TACTGAGG GAGGAGTC GGTACGTTG GCGGCGAAT ATGAGCTTA	300
AGTCTTCTG TACTGAGG GAGGAGTC GGTACGTTG GCGGCGAAT ATGAGCTTA	360

## 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 535 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

```

ACCGGCGCCA CCGGCGGCAC CGGGTTCGCC GGTGCGCCCG GCGGGGCCCC CGGGCAGGGC      60
GGTATCAGCG GTGCGGCGCG CACCAACGGC TCTGGTGGCG CTGGCGGCAC CGGCGGACAA      120
GGGCGGCGCG GGGGCGGTGG CGGGGCGCGC GCGGATAACC CCACCGGCAT CGGCGGCGCC      180
GGCGGCACCG CGCGCACCGG CGGAGCGGCG CGAGCGGCGG GGGCGGGTGG CCGCATCGGT      240
ACCGGCGCCA CCGGCGCGCG GGTGGGCAGC GTGCGTAACG CGGGGATCGG CGGTACCGCG      300
GGTACCGGTC GTGTGCGTGG TGCTGGTGGT GCAGGTGCGG CTGCGGCGCG TGGCAGCAGC      360
CGTACCGGTC GCGCGCGGTC CGCGGCGGCG GTGCGCGGCG AAGCGCGGAG CGGCGGCAAC      420
AGCGGTGTGG CGCGCACCAA CGGCTCGGCG CGCGCGGCGG GTGCAGGCGG CAAGGGCGGG      480
ACCGGAGGTG CGGCGCGGTC CGGCGCGGAC AATCGGCGCG GTGCTGGTTT CGCGG      535

```

## 2. INFORMATION FOR SEQ ID NO:167:

## 1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 590 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

## xi. SEQUENCE DESCRIPTION: SEQ ID NO:167:

```

CGGAGCTCGC CGGGCGCGATA CGGGGTCAC CGACTACTAC ATGATCGGCA CGGAGAATCG      60
GAGGCTGCTG GAGCGGCTTC CGGGCGGTCG GTGATGCGA GATGCTGTCG CGGAGCTGAT      120
GATGCTGCTG GAGCGGCTTC CGGGGTCG GTGATGCGA GATGCTGTCG CGGAGCTGAT      180
GATGCTGCTG GAGCGGCTTC CGGGGTCG GTGATGCGA GATGCTGTCG CGGAGCTGAT      240
GATGCTGCTG GAGCGGCTTC CGGGGTCG GTGATGCGA GATGCTGTCG CGGAGCTGAT      300
GATGCTGCTG GAGCGGCTTC CGGGGTCG GTGATGCGA GATGCTGTCG CGGAGCTGAT      360
GATGCTGCTG GAGCGGCTTC CGGGGTCG GTGATGCGA GATGCTGTCG CGGAGCTGAT      420

```

ATCAACGCGA TCGGCTATCC COTGGCGGCC ACCCTAGGTT TAGGCACGAT CGATAGCGGG 500  
 CGGCGTGGAA TTGCTCACCC TCCTCGCGGC GGCCTCGGAC ACCGTTGAA ACATCGAGGG 560  
 COTCGTCACC TAACGGATTC CGACGGCAT 590

## (2) INFORMATION FOR SEQ ID NO:168:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 407 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:168:

ACGGTACGG CGTACTGGG GCGGCGCAGC GCGGCAACGG CCGGAATCCC GGGTGGCTCT 60  
 TGGGCACAGC GGGGGGTGGC GGCAACGGTC GCGGCGCAGC CACCGGTACT GCAGGTGGCG 120  
 GGTCTGGGG CACCGGCGGC GAGGCGGGGA CCGGCGGCGC TGCGGCGCTG TTAATGGGCG 180  
 CCGGCGGCGC CCGGCGCGGT GCGACTGGCG GCGGCGGCGC TGCGGCTGTG GACCGTGGCG 240  
 GCGGCGGCGC GCGGCGGCGC GCGGCGGCGC ACAGCGGCGC CCGGCGTCAA GCGGCGCTGC 300  
 TGTTCGGCG CCGGCGCAGC GCGGAGCGC GCGGCTACGG CCGGCGATGGC GGTGGCGGCG 360  
 GTGACGGCTT CGACGCGAGC ATGCGCGGCT TGCTGGTAC GGTGGC 407

## (2) INFORMATION FOR SEQ ID NO:169:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 468 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:169:

ATGCTTAAI GCGATGCGC GCGGCGGCAA GCGATTCGCG GGTCTACCG CAGGACGTCG 60  
 GCGATTCGCG GCGGCGGCAA GCGGCGGCGC GGTCTACCG GGTCTACCG GGTCTACCG 120  
 ACAGCGAGTC GTGCGGCGCG GCGGAGGCGA ACAGCTCGGT GTGCGGCGTC GAGATCGGCG 180  
 GGTCTACCG GTGCGGCGAA GGTCTACCG GCGGCGGCGC GTGCTTGTGA TCGTGGAGCA 240  
 GTGCTTGTGA GTGCTTGTGA ACAGCGGCGC GCGGCGGCAA GGTGCGGCGC GTGCGGAGTA 300  
 GGTGCGGCAA GGTGCGGCAA GGTGCGGCGC GGTCTACCG GGTCTACCG GGTCTACCG 360

12) INFORMATION FOR SEQ ID NO:170:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

GGTGGTAACG GCGGCCAGGG TGGCATCGGC GCGCCCGCG AGAGAGCCGC CGACGGCGCC	60
GGCCCCAATG CTAACGGCGC AAACGGCGAG AACGGCGGTA GCGGTGCTAA CCGTGGCGAC	120
GGCGCGCGCG GCGGCAATGG CGGCGCGCGC GCGAACCGCG AGGCGGCGCG GTACACCGAC	180
GGCGCCACGG GCACCGCGCG GCACCGCGCG AACGGCGCG	219

13) INFORMATION FOR SEQ ID NO:171:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 494 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

TAGCTCCCGC GAGGCGCGCA AGGCGCGCGA CGGTGGCCAC GCGGCTGACG GCGTGGCGCG	60
CTACAGTTCC CTACCGGACG GCGGCGCGCG CGGTGGCGCG GCGGCGCGCG GCGGCGCGCG	120
CTACCGCGTTT TTGCGCGCGA AGGCGCGCGT TCGGCGCGAC GCGGCTGACG GCGGCGCGAA	180
GCGGCGCGCGT AGGCGCGCGA CGGTGGCGCG TCGGCGCGCG AACGGCGCGT TCGGCGCGCG	240
GCGGCGCGAC GCGGCTGTTT CGGTGGCGCG TCGGCGCGCG GCGGCTGCGT GCGGCGCGCG	300
CTATGCGCGC CTATGCGCGC GCGGCGCGCG TCGGCTGCGT GCGGCGCGCG CTGCGCGCGC	360
TAAGCGCGCG CTGCTGCGCG TAAGCGCGCG TACGCTGCGC AGGCGCGCGA AGTTCGCGCG	420
TAAGCGCGCG TCGGCTGCGT GCGGCGCGCG TCGGCGCGCG TCGGCGCGCG TCGGCGCGCG	480
AGTTCGCGCG GCGG	494

14) INFORMATION FOR SEQ ID NO:172:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 base pairs

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:172:

```

GGGCGGGTGG TGCCGCGGGC CAGCTCTTCA GCGCCGGAGG CGCGGCGGGT GCGGTGGGG      60
TTGCGCGCAC CGCGGCGCCAG GGTGGGGGTG GCGGTGCCGG AGCGGCGGGC GCGGACGCCC      120
CGCCGAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCGGGC GCGGTCCGGG      180
GCCGAGCGGG CAACGCCATT GCGGCGGGCA TCAACGGCTC      220

```

## (2) INFORMATION FOR SEQ ID NO:173:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 388 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:173:

```

ATGCGCGCAA CGGGGGCCCC GCGGTGCTG GCGGGGCCGG CGACTACAAT TTCCAACGGC      60
GGGCAGGCTG GTCCCGCGCG CCAAGGCGGC CAAGGCGGCG TCGCGGGGGC AAGCACCACC      120
TGATCGGCCT AGCCGCAACC GGGAAAGCCG ATCCAACAGG CGACGATGCC GCCTTCCTTG      180
CGGCGTTGSA CGAGGCGGCG ATCAGCTACG CTGACCCAGG CCACGCCATA ACGGCGGCCA      240
AGGCGATCTG TGGGCTGTGT GGTAAACGGC TAACAGGTCT ACAGCTGGTC GCGGACCTGC      300
GGGACTACAA TCGCGGGCTG ACCATGGACA GCGCGGCCCA GTTCGCTGCC ATGCGATCAG      360
TGGGCTACTG TCGCGAAGAC TTGGAACA      388

```

## (2) INFORMATION FOR SEQ ID NO:174:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:174:

```

TAAAGGCGCG TACCGGCGCG TCGCGGCTGA AGAGT TGA TCGGTGCTA TCGCGGCGAG      60
AGGCTGTCTA AGCGGCGACC CGCGGCGGCG GCGGCGAGGT CGGCGGCGCG GCGACCGGCT      120
TAAAGGCGAG CGCGGCGGCG AAGCGGCGCA AGCGGCTGTA TCGCGGCGGT TCGCGGCGAG      180
TAAAGGCGAG CGGAAAGGCT TAAAGGCGCA TAAAGGCGCT TCGCGGCGCG TAAAGGCGAG      240

```

(D) TOPOLOGY: linear

X1) SEQUENCE DESCRIPTION: SEQ ID NO:177:

```

AGCAGCGCTA CCGGTGGCGC CGGGTTCGCC GGGGGCGCGG GCGGAGAAGG CGGAGCGGGC      60
GGCAACAGCG GTGTGGGCGG CACCAACGGC TCCGGCGGCG CCGGCGGTGC AGGCGGCAAG      120
GGCGGCACCG GAGGTGCCGG CGGGTCCGGC GCGGACAACC CCACCGGTGC TGGTTTCGCC      180
GGTGGCGCGG GCGGCACAGG TGGCGCGGCC GCGCGCGGCG GGGCGGCGCG GGCAGCCGGT      240
ACCGCGCGCA CCGGCGGCGT TGTGCGGCCG ACCGGTAGTG CAGGCATCGG CCGGCGCGGC      300
GGCGCGCGCG GTGACGGCGG CGATGGGGCG AGCGGTGTGG GCCTGGGCGT CTCGGGCTTT      360
GACGGCGCGG AAGCGCGCGA AGGCGGCGCG GCGGCGCGCG CCGGCGCGCG CGGCATCAAC      420
GGGCGCGCGG GGGCGGCGCG CAACCGCGCG GACGCGCGCG ACGGCGCAAC CGGTGCGCGA      480
GTTGTGCGCG ACAACGCGCG GTTGGCGCGT GACGCTGGCG CCGGTGGCGG CGCGGCGCAAC      540
GGGCGCAACG CCGGCGTGGG CCGTACAGCG AAAGCGCGCG AGGGCGGCGG CCGGCGCAAT      600
GGGCGCAACG GGGCGGCGCG CGGTGTGGCG GCGGCGCGCG ACAACAATTT CAACGGCGCG      660
GAGGTGTGTT CCGGCGCGCA AGGCGGCGCA GCGGCGTTGG CCGGCGCGCA CACCACCTGA      720
TGTGTGTAGG CGCACCTGGG AAAGCGGATG GACAGGCGGA CGATGCGCGG TTCCTTGCGG      780
GTTGTGAGCA GCGGCGGATG AGGTAGGCTG AGGCGGCGCA GCGGCGGATG GCGGCGGATG      840
GATGTGTGTT GGTGTGTGTT AGGTGTGTAA GAGGTGTGTA GTGTGTGTTG GACCTGCGCG      900
GATAGATATG CCGGCTGAGG ATGGATAGCG GCGGCGGATG GGTGTGTGTT GATCAGGCG      960
GTACTGCGCG CGAACGCTG GAACA

```

X2) INFORMATION FOR SEQ ID NO:178:

X2) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1134 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

X3) SEQUENCE DESCRIPTION: SEQ ID NO:178:

```

GATGATGAGG ATGGGTAGCG GCGGCGGATG GAGGTGTGTT GTTGTGTGTT GTTGTGTGTT      60

```

GGGGGACGG TGCACCTCTCA GGCAGCACCG GTGGTGCCGG

400

(2) INFORMATION FOR SEQ ID NO:175:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 538 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

GGCAACGGCG GCAACGGCG CATCGCCGGC ATTGGGCGG AACGGCGTTC CGGGACGGGC	60
AGCGGCAACG GCGGCAACG GCGGCAGCGG CGGCAACGGG GGCAACGGCG GCATGGGCGG	120
CAACAGCGGC ACCGGCAGCG GCGACGGCGG TCGCGCGCGG AACGGCGCGG CGGCGGGCAC	180
GGGCGGCACC GCGGCGGACG GCGGCGTCA GGTACTGGC GGCACCGCGG GCAGCGGTGG	240
CACCGCGCGT GACGCGCGTA ACGGCGGCAA GCGAGCAGAT AACACCGCAA ACATGACTGC	300
GCAGCGCGCG GGTGACGGTG GCAACGGCGG GACCGGTGGC TTGGGCGCGG GGGCCCGGGC	360
CGCGCGCGGT GCGTTGACCG CTGGCGGCAA GCGCACCGCG GGGCAAGCGG GCGCCCGCGG	420
CGATGGCGCG AACGGGCGCA TCGGCGGCGA GGGCGGCTC ACTGACGACC CCGGCGGCAA	480
GGGGGCGACC GCGGCGAACC GCGGCACCGG GCGCACCGCG GCGCGCGGCA TCGGCAGC	538

(2) INFORMATION FOR SEQ ID NO:176:

(1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 339 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:176:

GGGCGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	60
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	120
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	180
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	240
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	300
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	360
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	420
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	480
GGGCGGCGG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	540

(2) INFORMATION FOR SEQ ID NO:177:

AGATTTTGGAA CAGGGCCCAAC GAGGTGGAGG CCCCAGTGGC GGACCCACCG ACTGATGTCC 300  
 CCATCAGACC GTGCGAACTC ACCGGGGCTA AAAACGCCGC CCAACAGCTG GTATTGTCCG 360  
 CCGACAACAT GCGGGAATAC CTGGCGGGCG GTGCCAAAGA GCGGCAGCGT CTGGCGACCT 420  
 CGGTGCGCAA CGCGGCCAAG GCGTATGGCG AGGTTGATGA GGAGGCTGCG ACCGCGCTGG 480  
 ACAACGACCG CGAAGGAACT GTGCAGGCAG AATCGGCCGG GCCCGTCGGA GGGGACAGTT 540  
 CGGCGGAACT AACCGATACG CCGAGGGTGG CCACGGCCGG TGAACCCAAC TTCATGGATC 600  
 TCAAGAAGC GGCAAGGAAG CTCGAAACCG GCGACCAAGG CGCATCCCTC GCGCACTTTG 660  
 CGGATGGGTG GAACACTTTT AACCTGACCG TGCAGGCGA CGTCAAGCGG TTCGGGGGT 720  
 TTGACAACTG GGAAGGGGAT CGGGGTACCG GTTCGGAGGC TTGGCTCGAT CAACAACCGC 780  
 AATGGATACT CCACATGGCC AAATTGAGCG CTGGGATGGC CAAGCAGGCT CAATATCTCG 840  
 CGCAGCTGCA CGTGTGGGCT AGGGGGGAAC ATCGGACTTA TGAAGACATA CTGGGGCTCG 900  
 AACGGCTTTA CGCGGAAAAC CTTTCGGGCC GCGACCAAT TCTCCCGGTG TACCGCGAGT 960  
 ATCAGCAGAG CTCGAGAGAG GTGCTGACCG TATCAACAA CAAGGCAGCC CTGGAACCGC 1020  
 TAAAGCCCGC GAAGCGTCCG CCGCCCATCA AGATCGACCG GCGCCCGCTT CCGCAAGAGC 1080  
 AGCGATTGAT CGCTGGCTTC CTGATGCCCG CGTCTGACCG CTGCGGTGTG ACTCCCGGTA 1140  
 CGCGGATCGC AGCGGCACCG ATGGTTCCCG CTACCGGATG CGCGGCTGCT TGGCTCCCGC 1200  
 CTGACACCGC CGCCGACCTG ACCTCGGCTG CGCGGGAAGG CGCAGGCTG TCGCGGAGC 1260  
 TCGCGGTGCA AGCGGCATCG CTGGGTGGCG CTGGAGGCGG CGGGGTGGCG TCGCGCGCT 1320  
 TCGGATCGCG CATCGCGGCG CGCGAATCGG TCGCGCGCGG TCGCGCTGCT GACATTGCGG 1380  
 ATGAGTCTG ATGAGGCGCG CGCGCGCGCG CTGCGTCTCG CGCGGCTGCT ATGGAATCT 1440  
 ATGAGTCTG ATGAGGCGCG CGCGCGCGCG CTGCGTCTCG CGCGGCTGCT ATGGAATCT 1500  
 ATGAGTCTG ATGAGGCGCG CGCGCGCGCG CTGCGTCTCG CGCGGCTGCT ATGGAATCT 1560  
 ATGAGTCTG ATGAGGCGCG CGCGCGCGCG CTGCGTCTCG CGCGGCTGCT ATGGAATCT 1620  
 ATGAGTCTG ATGAGGCGCG CGCGCGCGCG CTGCGTCTCG CGCGGCTGCT ATGGAATCT 1680  
 ATGAGTCTG ATGAGGCGCG CGCGCGCGCG CTGCGTCTCG CGCGGCTGCT ATGGAATCT 1740

CCGCGCGGGC GAGCAGCTGA CCGCTGCGTT ATCGGCCATG TCCCGCGCGA TGAACGAAGG 1920  
 AATGGCCTAA GCCCATTGTT GCGGTGGTAG CGACTACGCA CCGAATGAGC GCCGCAATGC 1980  
 GGTCAATCAG CGCGCCCGAC ACGGCGTGAG TACGCAATTGT CAATGTTTTG ACATGGATCG 2040  
 GCGGGGTTCTG GAGGGCGCCA TAGTCCTGGT CGCCAATATT GCCGCAGCTA GCTGGTCTTA 2100  
 GGTTCGGTTA CGCTGGTTAA TTATGACGTC CGTTACCA 2138

## (2) INFORMATION FOR SEQ ID NO:179:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 460 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (2) SEQUENCE DESCRIPTION: SEQ ID NO:179:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn  
 1 5 10 15  
 Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val  
 20 25 30  
 Pro Ile Thr Pro Cys Glu Leu Thr Ala Ala Lys Asn Ala Ala Gln Gln  
 35 40 45  
 Leu Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala  
 50 55 60  
 Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala  
 65 70 75 80  
 Tyr Gly Glu Val Asp Glu Gln Ala Ala Thr Ala Leu Asp Asn Asp Gly  
 85 90 95  
 His Gly Thr Val Gln Ala His Ser Ala Thr Ala Val His Gly Asn Ser  
 100 105 110 115  
 Pro Ala Ala Leu Thr Asn Thr Pro Arg Val Ala Thr Ala Gly Ala Ser  
 120 125  
 Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Cys Leu Glu Thr Gly Asp  
 130 135 140  
 Thr Gly Ala Ser Leu Ala His Phe Ala Asp Glu Trp Asn Thr Phe Asn  
 145 150 155 160

Gln Trp Ile Leu His Met Ala Lys Leu Ser Ala Ala Met Ala Lys Gln  
 195 200 205  
 Ala Gln Tyr Val Ala Gln Leu His Val Trp Ala Arg Arg Glu His Pro  
 210 215 220  
 Thr Tyr Glu Asp Ile Val Gly Leu Glu Arg Leu Tyr Ala Glu Asn Pro  
 225 230 235 240  
 Ser Ala Arg Asp Gln Ile Leu Pro Val Tyr Ala Glu Tyr Gln Gln Arg  
 245 250 255  
 Ser Glu Lys Val Leu Thr Glu Tyr Asn Asn Lys Ala Ala Leu Glu Pro  
 260 265 270  
 Val Asn Pro Pro Lys Pro Pro Pro Ala Ile Lys Ile Asp Pro Pro Pro  
 275 280 285  
 Pro Pro Gln Glu Gln Gly Leu Ile Pro Gly Phe Leu Met Pro Pro Ser  
 290 295 300  
 Asp Gly Ser Gly Val Thr Pro Gly Thr Gly Met Pro Ala Ala Pro Met  
 305 310 315 320  
 Val Pro Pro Thr Gly Ser Pro Gly Gly Gly Leu Pro Ala Asp Thr Ala  
 325 330 335  
 Ala Gln Leu Thr Ser Ala Gly Arg Glu Ala Ala Ala Leu Ser Gly Asp  
 340 345 350  
 Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Val  
 355 360 365  
 Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg  
 370 375 380  
 Pro Ala Gly Ala Gly Asp Ile Ala Glu Leu Gly Gln Gly Arg Ala Gly  
 385 390 395 400  
 Ile Gly Ala Ala Leu Ile Ile Ile Val Ile Val Pro Val Ile Ala  
 405 410 415  
 Ala His Gln Gly Gln Gly Gly Ala Lys Ser Lys Gly Ser Gln Gln Glu  
 420 425 430  
 Asp Glu Ala Leu Tyr Thr Glu Asp Arg Ala Trp Thr Glu Ala Val Ile  
 435 440 445  
 Ser Asn Arg Arg Arg Gln Asp Ile Ile Ile Ile Ile Ile Ile Ile

- (A) LENGTH: 277 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

Ala Gly Asn Val Thr Ser Ala Ser Gly Pro His Arg Phe Gly Ala Pro  
 1 5 10 15  
 Asp Arg Gly Ser Gln Arg Arg Arg Arg His Pro Ala Ala Ser Thr Ala  
 20 25 30  
 Thr Glu Arg Cys Arg Phe Asp Arg His Val Ala Arg Gln Arg Cys Gly  
 35 40 45  
 Phe Pro Pro Ser Arg Arg Gln Leu Arg Arg Arg Val Ser Arg Glu Ala  
 50 55 60  
 Thr Thr Arg Arg Ser Gly Arg Arg Asn His Arg Cys Gly Trp His Pro  
 65 70 75 80  
 Gly Thr Gly Ser His Thr Gly Ala Val Arg Arg Arg His Gln Glu Ala  
 85 90 95  
 Arg Asp Gln Ser Leu Leu Leu Arg Arg Arg Gly Arg Val Asp Leu Asp  
 100 105 110  
 Gly Gly Gly Arg Leu Arg Arg Val Tyr Arg Phe Gln Gly Cys Leu Val  
 115 120 125  
 Val Val Phe Gly Gln His Leu Leu Arg Pro Leu Leu Ile Leu Arg Val  
 130 135 140  
 His Arg Glu Asn Leu Val Ala Gly Arg Arg Val Phe Arg Val Lys Trp  
 145 150 155 160  
 Phe Glu Pro Asp Tyr Val Phe Ile Ser Arg Met Phe Pro Pro Ser Pro  
 165 170 175  
 His Val Gln Leu Arg Arg His Leu Ser Leu Leu Ile His Arg Ser Ala  
 180 185 190  
 Gln Phe Gly His Val Gln Tyr Pro Leu Ser Leu Leu Ile Gln Arg Ser  
 195 200 205  
 Leu Ala Ser Gly Ser Arg Ile Ala Phe Pro Val Val Lys Pro Pro Glu  
 210 215 220  
 Pro Leu Asp Val Ala Leu Gln Arg Gln Val Gln Ser Val Pro Pro Ile  
 225 230 235

Cys Arg Phe Phe Glu Ile His Glu Val Gly Phe Thr Gly Arg Gly His  
 260 265 270

Pro Arg Arg Ile Gly  
 275

(2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:181:

Ala Val Ala Ala Ser Phe Ile Asp Trp Leu Asp Ser Pro Asp Ser Pro  
 1 5 10 15  
 Leu Asp Pro Ser Leu Val Ser Ser Leu Leu Asn Ala Val Ser Cys Gly  
 20 25 30  
 Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg  
 35 40 45  
 Trp Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr  
 50 55 60  
 Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg  
 65 70 75 80  
 Ser Val Ala Pro Arg Ala Glu Phe Gly Thr Arg Leu Ala Ser His Arg  
 85 90 95  
 Ala Ser Pro Ser Asn Ala Cys Pro Val Arg Ile Val Thr Ser Ala Ser  
 100 105 110  
 Gly Arg Pro Ile Ser Ser Pro Pro Ile Val Arg Ser Arg Ser Cys Val  
 115 120 125  
 Asn Asp Asn Ile Arg Arg Cys Ala Ser Ile Tyr Arg Arg Leu Asn Arg  
 130 135 140  
 Ala Arg Ser Ser Ser Ser Ala Arg Arg Cys Arg Thr Ile Gly Thr Phe  
 145 150 155 160  
 Arg Arg Ser Arg Thr Ser Ala Ser Met Arg Val Ser Thr Asn Ser Pro  
 165 170 175  
 Ser Val Thr Ala Glu Val Ala Ser Gly Val Thr Arg Ala Thr Ser Gly

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 196 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

Gln Glu Arg Pro Gln Met Cys Gln Arg Val Ser Glu Ile Glu Pro Arg  
 1 5 10 15  
 Thr Gln Phe Phe Asn Arg Cys Ala Leu Pro His Tyr Trp His Phe Pro  
 20 25 30  
 Ala Val Ala Val Phe Ser Lys His Ala Ser Leu Asp Glu Leu Ala Pro  
 35 40 45  
 Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp Ala Glu Asp Arg Arg Val  
 50 55 60  
 Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala  
 65 70 75 80  
 Gly Gly Glu Ala Asp Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Gln  
 85 90 95  
 Ala Gly Arg Val Arg Arg Gly Gln Glu Leu His His Arg His Arg His  
 100 105 110  
 Gln Gly Ala Ala Pro Asp Leu Arg Arg Arg Arg Arg His Arg Arg Val  
 115 120 125  
 Gln His His Arg Arg Leu Gln Arg Val Arg Gln Leu Arg Arg Tyr Val  
 130 135 140  
 Gln Thr Ala His His Arg Arg Phe Ala Arg Thr Asp Arg Val Arg His  
 145 150 155 160  
 His Val Arg Gly Pro Ser Asn His Arg Arg Arg Arg Val Thr Arg Gly  
 165 170 175  
 Arg His Ser His Ala Gly His His Arg Val Arg His Ala His Ser Val  
 180 185 190  
 Gly Gly Ser Ala  
 195

## 2. INFORMATION FOR SEQ ID NO:183:

## SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 196 amino acids

X1 SEQUENCE DESCRIPTION: SEQ ID NO:183.

```

Val Arg Cys Gly Thr Leu Val Pro Val Pro Met Val Glu Phe Leu Thr
1           5           10           15

Ser Thr Asn Ala Pro Ser Leu Pro Ser Ala Tyr Ala Glu Val Asp Lys
20           25           30

Leu Ile Gly Leu Pro Ala Gly Thr Ala Lys Arg Trp Ile Asn Gly Tyr
35           40           45

Glu Arg Gly Gly Lys Asp His Pro Pro Ile Leu Arg Val Thr Pro Gly
50           55           60

Ala Thr Pro Trp Val Thr Trp Gly Glu Phe Val Glu Thr Arg Met Leu
65           70           75           80

Ala Glu Tyr Arg Asp Arg Arg Lys Val Pro Ile Val Arg Glu Arg Ala
85           90           95

Ala Ile Glu Glu Leu Arg Ala Arg Phe Asn Leu Arg Tyr Pro Leu Ala
100          105          110

His Leu Arg Pro Phe Leu Ser Thr His Glu Arg Asp Leu Thr Met Gly
115          120          125

Gly Glu Glu Ile Gly Leu Pro Asp Ala Glu Val Thr Ile Arg Thr Gly
130          135          140

Tln Ala Leu Leu Gly Asp Ala Arg Trp Leu Ala Ser Leu Val Pro Asn
145          150          155          160

Ser Ala Arg Gly Ala Thr Leu Arg Arg Ser Gly Ile Thr Asp Val Ala
165          170          175

Asp Leu Arg Ser Ser Arg Glu Val Ala Arg Arg Gly Pro Gly Arg Val
180          185          190

Pro Asn Gly Ile Asp Val His Leu Arg Trp Phe Pro Asp Leu Ala Asn
195          200          205

Asp Asp Ala Arg Asp Ser Ala Arg His Glu Thr Ala Thr Lys Arg Leu
210          215          220

Glu Thr Asn Asp Gly Ser Asn Gly Gly Ser Gly Glu Ser Ser Glu Ser
225          230          235          240

Ile Asn Asp Ala Ala Thr Arg Tyr Met Thr Asp Glu Thr Arg Glu Phe
245          250          255

Ser Thr Arg Asn Ile Ala Thr Arg Arg Ala Thr Thr Thr Thr Thr

```

Arg Thr Gly Phe Val Val Ala Leu Val Leu Glu Ala Val Gly Leu Asp  
 290 295 300

Arg Asp Val Ile Val Ala Asp  
 305 310

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2072 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:184:

TTGCTGCGCA TTCCGCACGA GCTCAGCAGC CCAAGGGGCG GTTGGGGGAA GTCATCGAGG 50  
 CATTGCGGCA GGGGCTCGCC GCGAAGGGTA AGCAAATCAA CACCACGCTG AACAGCCTGT 120  
 CCGAGGCGTT GAAGCGCTTG AATGAGGGCG GCGCGGACTT GTTGGCGGTG GTACCGACCG 180  
 TCGCGGTATT CCGCAACGCG CTACATCAGG AGGACCAACA GTTCGTGCGG TTGAACAAGA 240  
 AGCTTGGCGA GTTCACCGAC AGGTTGACCC ACTCGGATCG GGAGCTGTCT AACCGCATCG 300  
 AGCAATTGCA CAGCTTCTCT GCGCTCGCCG GCGGCTTCTT GCGCAAGAAG CCGGAGGTGC 360  
 TGAGCGATGA CCGCAATAAT CTGCGGACCG TGACCACGAC GTTCTGTGAG CCGGATCGGT 420  
 TCGATCGGTT CGAGACGCTG CTGCACATCT TCGCGACGCT GCGCGCGAAT ATTAAGCAGC 480  
 TTTAGCATCG CAGACAGGCT CGGCTGCTGT GCGTTGCGCG GTTCAGGAAT TTGGGCAAGC 540  
 TGATGCACTT CATCTGCACT TCGAATCAGG TCGGTAGCCG GTCGCGTTAT CAAGAGCTCG 600  
 CGGAAGCTCT TCGCGCATAT TTGCGGCGAG TCGTGGATCG GATCAAGTTC AACTAGTTTC 660  
 GTTCTGCTT GAAGCTGCTT AGGAGGCTT CGACCTGCTT CAAGAGATG TGTAGTCTG 720  
 ATGCTGCTT CGAGCGCGCG AAATCGGAGA CGGACAGGAT GTGCTGCTT ATCTGCTT 780  
 ATGATAGCG GTTCTGCACT TCGAATCAGG TCGGTAGCCG GTTCAGGAAT TTGGGCAAGC 840  
 CGGAGCTTGA CTGCGGACCG ATCAGCGAGG GTTCTGCTGAG CCGGAGTCTG CTGCGGCAAG 900  
 TATGCTGCTG TCGCGCATAT GCGGCTGCTT TGTCAAGGCT CCAAGACCGCG CCGGAGCGCG 960  
 ATGATAGCTT CAGCGCATAT GCGGCTGCTT TGTCAAGGCT CCAAGACCGCG CCGGAGCGCG 1020

TCTTGTTCCT GTGCGCGGGC GTGCGGACCT TCCTGTTCGG GGTGTCTCTT AGCCCGCGCC	1200
GTGGAACGAT GCGCGATCGG CACGTCTTGA TACCGGCGAT CACCGGCGTG GCCTTGATCG	1260
CGGCATTGCT CGCACATTCT TGSTACCGCA CAGAACATCC GCTCATAGAC ATGCGGCTTGT	1320
TCCAGAACCG AGCGGTCCGG CAGGCCAACA TGACGATGAC GGTGCTCTCC CTCGGGCTGT	1380
TGGGCTCCTT CTTGCTGCTC CCGAGCTACC TCCAGCAAGT GTTGCACCAA TCACCGATGC	1440
AATCGGGGGT GCATATCATC CCACAGGGCC TCGGTGCCAT GCTGGCGATG CCGATCGCCG	1500
GAGCGATGAT GGACCGACCG GGACCGGCCA AGATCGTGCT GGTGGGGATC ATGCTGATCG	1560
CTGCGGGGTT GGGCACCTTC GCTTTTGGTG TCGCGCGGCA AGCGGACTAC TTACCCATTC	1620
TGCGGACCGG GCTGGCAATC ATGGGCATGG GCATGGGCTG CTCGATGATG CCACTGTCCG	1680
GGCGCGGAGT GCAGACCGTG GCGGCACATC AGATCGCTCG GGTTCGAGC CTGATCGGCG	1740
TCAACGAGCA GGTGGGCGGT TGATAGGGA CCGCACTGAT GTGGGTGCTG CTCACCTACC	1800
AGTTCAATCA CAGCGAAATC ATCGCTACTG CAAAGAAAGT CGCACTGACC CCAGAGAGTG	1860
CGCGCGGGCG GGGGGCGGCG GTTGACCTTT CTTGCTTACC GCGCCAAAGC AACTTGGGCG	1920
CGCAACTGCT GCATGACCTT TCGCAGGCT AGCGGTGGT ATTGGTGATA GCGACCGGCG	1980
TAGTGTCTG CAGCGTATC CCGCGGCGAT TCCTGCCGAA ACAGCAGGCT AGTCATCGAA	2040
AGCGAGGCTT GCTATCGGCA TGACCTCTCG TT	2070

# 2. INFORMATION FOR SEQ ID NO.135.

## A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1903 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## B. SEQUENCE DESCRIPTION: SEQ ID NO.135.

TTTCTTCTTA TAACTCTTTT TTGGA TAACT TGAATGCA TTCTCTTCTT ATCTTCTTAA	1
TTCTCTTCTT TATTCTTAA TAACTCTTCT TCAATCTCTT TCACTCTCTT CTCTCTCTCT	120
TTCTCTCTCT CTCTCTCTCT TTCTCTCTCT TCACTCTCTT TCACTCTCTT AACTCTCTCT	180
TTCTCTCTCT CTCTCTCTCT TCACTCTCTT TCACTCTCTT TCACTCTCTT TCACTCTCTT	240
TTCTCTCTCT CTCTCTCTCT TCACTCTCTT TCACTCTCTT TCACTCTCTT TCACTCTCTT	300

GAGTTGGCGG	GGGCGGAATT	GGGGCATTGC	GTGGAAGGCG	AGCGGATCCC	GGCGCGCGCG	480
CGCGGTGGCT	GGTGTTTTGG	GGCGCGGGAT	GGCCACGACG	AGAAAGACGA	TGGCGCGGAT	540
GAACAGCGCC	ACGGCAATCA	CGACCAGCAG	ATTTCCGACG	CATACCGCTC	CGTACCGGTG	600
CGCGCGGGTT	GGTCGATCGG	TGGCATATCG	ATGGCGCCGT	TTAACGTAAC	AGCTTTGCGG	660
GGACCGGGGG	TCACAACGGG	CGAGTTGTCC	GGCGGGGAAC	CCGGCAGGTC	TGGCGCGCGG	720
TCACCCGAGC	TCACTGGTGC	ACCATCCGGG	TGTCGGTGAG	CGTGCAACTC	AAACACACTC	780
AACGCGAAG	GTTCCTCAGG	TCACCAGCTC	AACCTGACG	CGCAATCGCT	CGTACGTTTC	840
TACGCGCGCG	AGGTGCGGAG	TCAGCAGCTT	TGCGCGCGCA	GCTTTGCGCG	TGAAGCGGAC	900
TAGGCGATCG	TAGGTTGCGC	CACCGGTGAC	ATCGTGCTCG	GCGAGGTGGT	CGGTCAAGCG	960
CGATATGAG	CAGGCATCCA	GTGCCAGGTA	GTGCTGGAG	GTGATGTCCG	CGAGTAGGGC	1020
GTGACCGCGA	ACAGGGCGAA	TACGATGCGG	CGGTGCTAGC	CGGTCAAGA	CGCAATAGGT	1080
TTGACAGCGC	GGGTGCGCGA	TCAGATGGAC	GCGACGCTTC	AGCGCGCGCA	CGGCGCGCTC	1140
GTGCGCTTCG	TGCGAGCTCG	CGAATCCGCG	AACGACGACG	GTGGTGTCTG	GTGCGATCAC	1200
CGCGCTGTCC	GATCGAGCGT	TTCCCGAAG	ATTTGCTCGG	TGAACCGCGG	CAGGGGACGT	1260
TGTGCGCGTG	CGACGAGAAC	CGAGCGCTTC	CGAAGGAGTT	CGACACCGGT	CGGGGCGGGC	1320
TGAATGTGCA	TGCGCGCATC	CGCGCTCGTG	ATGTCGACGT	GTCGCTTCGG	CGCGAGCGCA	1380
CGCGCGCTCG	GAATCGCGCT	GGAATCAGC	AGACGCTGCT	CGACATCGAT	GCTTGTTCGG	1440
ATGCTAGGAA	ATTTAGCATC	CGACGTTCCA	TAGGCGTCTC	GTGCGCGCGA	TGTGCGGACG	1500
ATCGCGTAGC	GTATCGAAGC	ATTGTTTTCG	AAATGCGCTG	CGGAGCGCTC	GCTGCGGGTG	1560
CGCGCTGCTC	ATGCGCGGTT	GACCGGATCT	GTGCTGCTCG	CGATCGGAGC	CGCGCTGCTC	1620
CGCGCTGCTC	GTGCTGCTCG	GACCGGATCT	GTGCTGCTCG	CGATCGGAGC	CGCGCTGCTC	1680
CGCGCTGCTC	GTGCTGCTCG	GACCGGATCT	GTGCTGCTCG	CGATCGGAGC	CGCGCTGCTC	1740
CGCGCTGCTC	GTGCTGCTCG	GACCGGATCT	GTGCTGCTCG	CGATCGGAGC	CGCGCTGCTC	1800
CGCGCTGCTC	GTGCTGCTCG	GACCGGATCT	GTGCTGCTCG	CGATCGGAGC	CGCGCTGCTC	1860
CGCGCTGCTC	GTGCTGCTCG	GACCGGATCT	GTGCTGCTCG	CGATCGGAGC	CGCGCTGCTC	1920

## 1 SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1055 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:186:

```

CTGGCGTCCC AGTGTACCCG GCGATATGAC GTCGGCATTG AATTTGCGGG CCCC GCCCGGA      60
CCCGTCGCCA CCCAATCTGG ACCACCCGGT CCGTCAATTG CCGAAGGTCG CCAAGTCCGT      120
GCCCCAATG TGCTGGGTT TCTTGAACGA AGGCCTGCCG TATCGGGTGC CCTACCCCCA      180
AACAACGCCA GTCCAGGAAT CCGGTCCCGG GCGGCCGATT CCCAGCGGCA TGTGCTAGCC      240
AGGCTGCTT GAGACCTAAC GGTGCTTAG GTGAAACCG GCGCCAGGGC CGCTGGACGG      300
GCTCATGCCA GCGAAATTAG AAAACCCCGG ATATTGTCCG CGGATTGTCA TAGGATGCTG      360
ATGCTTCTGT GGTTCGTGTT TAGCCATTGA GTGTGGATGT GTTGAGACCC TGGCCTGGAA      420
GGGACAACG TGCTTTTCCC TCTTGGTCCG CCTTTGCCCG CCGACGCGGT GGTGGCGAAA      480
GCGCTGAGT CCGGAATGCT GCGCGGCTTG TCGCTTCCCG TCAGCTGGGG AGTGGCTGTG      540
GCACCCGATG ATTATGACCA CTGGGCGGCT GCGCCCGAAG ACAGCGCCGA TGTGATGTC      600
GAGCGCGCCG AAGGGGCGGA GCGAGAGGCG GCGGCCATCG ACCAGTGGGA TGAGTGGCAG      660
GCTCGAAGC AGTGGGTGGC TGAGAACTGT GAACCCCGCT CTCAGGTGCC ACCGAGTAGC      720
AGCAGCTGTA TTCCGCTTC TCGGCTGCTT GGTAGAGA GCGGCGCCAG ACTGTCTTA      780
CTTGACCACT GATCGGCGCT CTGGTCTTT GCGGCGCGG CTATGACAGC AGTCATCTG      840
GATGACCACT TACAGGTATT AGGTCCAGGT TCAACAAGCA GACAGGCAAC ATGGCAACAC      900
GTTTATGAG GATGCGCA TCGATGCGA GAGGCTGCTT GGTGCTGAG ATGCAAGCG      960
GAGGCTGAG GAGGAGCTT GCGGCTGCTT GAGGCTGCTT GAGGCTGCTT GAGGCTGCTT      1020
GAGGCTGCTT GAGGCTGCTT GAGGCTGCTT GAGGCTGCTT GAGGCTGCTT GAGGCTGCTT      1080

```

2 INFORMATION FOR SEQ ID NO:187:

## 1 SEQUENCE CHARACTERISTICS

- A LENGTH: 1055 base pairs  
 B TYPE: nucleic acid  
 C STRANDEDNESS: single  
 D TOPOLOGY: linear

TCCGGGCTGA CCACCGGGAT CCGCGAACCA TCCGAGATCA CCTCGCAATG ATCCACCTCG	120
CGCAGCTGGT CACCCAGCCA CCGGGCGGTG TCGCAGACCG CCTGCATCAC CTTGGTATAG	180
CCCTCGCGCC CCAGCCGCAG GAAGTTGTAG TACTGGCCCA CCACCTGGTT ACCGGGACGG	240
GAGAAGTTCA GGGTGAAGGT CGGCATGTCG CCGCCGAGGT AGTTGACCCG GAAAACCAGA	300
TCCTCCGGCA GGTGCTCGGG CCCGCGCCAC ACGACAAACC CGACGCCGGG ATAGGTCAG	359

## (2) INFORMATION FOR SEQ ID NO:188:

## 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## X1. SEQUENCE DESCRIPTION: SEQ ID NO:188:

AACGGGCCCC TGGCACCCG TCGTSTAAGG GGTGTGTTT GTCCCATGAA GTGCTGAAC	60
GATGCATCTT GGCAGATTCC CGCCAGAGCA AAACAGCCCG TAGTCTAGT CCGAGTCGCC	120
CCCAAAGTTC CTCGAATAAC TCGTACCCG GAGGSCCAA CCGGCTCTCG TTCGCTAAGC	180
TGCGCGAACC ACTTGAGGTT CCGGCACTCC TTGAGCTCCA GACCGATTCC TTCGAGTGGC	240
TCATCGCTTC CCGGCGCTCG CGCGAATCCG CCGCGAGCG GGTGATGTC AACCCAGTGG	300
TTGCGCTCGA AGAGGTGCTC TACGAGCTGT CTGCGATCGA GCACTTCTCC	359

## 2. INFORMATION FOR SEQ ID NO:189:

## 1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 672 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

## X1. SEQUENCE DESCRIPTION: SEQ ID NO:189:

Met Ile Thr Lys Lys Ile Thr Gly Ala Val Ile Glu Ala Ser Ala Asp	1	10	15
Ile Leu Ala Gly Lys Gly Lys Glu Ile Asn Thr Thr Ser Asn Ser Leu	20	25	30
Met Ile Glu Lys Asn Ala Leu Asn Ile Glu Arg Gly Asn Phe Phe Ala	35	40	45

Gln Gln Phe Val Ala Leu Asn Lys Asn Leu Ala Glu Phe Thr Asp Arg  
 65 70 75 80  
 Leu Thr His Ser Asp Ala Asp Leu Ser Asn Ala Ile Gln Gln Phe Asp  
 85 90 95  
 Ser Leu Leu Ala Val Ala Arg Pro Phe Phe Ala Lys Asn Arg Glu Val  
 100 105 110  
 Leu Thr His Asp Val Asn Asn Leu Ala Thr Val Thr Thr Leu Leu  
 115 120 125  
 Gln Pro Asp Pro Leu Asp Gly Leu Glu Thr Val Leu His Ile Phe Pro  
 130 135 140  
 Thr Leu Ala Ala Asn Ile Asn Gln Leu Tyr His Pro Thr His Gly Gly  
 145 150 155 160  
 Val Val Ser Leu Ser Ala Phe Thr Asn Phe Ala Asn Pro Met Glu Phe  
 165 170 175  
 Ile Cys Ser Ser Ile Gln Ala Gly Ser Arg Leu Gly Tyr Gln Gln Ser  
 180 185 190  
 Ala Glu Leu Cys Ala Gln Tyr Leu Ala Pro Val Leu Asp Ala Ile Lys  
 195 200 205  
 Phe Asn Tyr Phe Pro Phe Gly Leu Asn Val Ala Ser Thr Ala Ser Thr  
 210 215 220  
 Leu Pro Lys Gln Ile Ala Tyr Ser Gln Pro Arg Leu Gln Pro Pro Asn  
 225 230 235 240  
 Gly Tyr Lys Asp Thr Thr Val Pro Gly Ile Trp Val Pro Asp Thr Pro  
 245 250 255  
 Leu Ser His Arg Asn Thr Gln Pro Gly Trp Val Val Ala Pro Gly Met  
 260 265 270  
 Ala Gln Val Gln Val His Pro Ile Thr Gln His Leu Leu Thr Pro Gln  
 275 280 285  
 Ser Leu Ala His Leu Met Gln Gln Pro Asn Ile Ala Pro Ser Ser Ser  
 290 295 300  
 Gly Leu Gln Thr Pro Pro Gly Pro Pro Asn Ala Tyr Asp Gln Tyr Pro  
 305 310 315 320  
 Val Leu Pro Pro Ile Gln Leu Gln Ala Pro Gln Val Pro Ile Pro Pro  
 325 330 335

355	360	365
Asp Tyr Met Gly Leu Leu Leu Leu Ser Pro Gly Leu Ala Thr Phe Leu 370 375 380		
Phe Gly Val Ser Ser Ser Pro Ala Arg Gly Thr Met Ala Asp Arg His 385 390 395 400		
Val Leu Ile Pro Ala Ile Thr Gly Leu Ala Leu Ile Ala Ala Phe Val 405 410 415		
Ala His Ser Trp Tyr Arg Thr Glu His Pro Leu Ile Asp Met Arg Leu 420 425 430		
Phe Gln Asn Arg Ala Val Ala Gln Ala Asn Met Thr Met Thr Val Leu 435 440 445		
Ser Leu Gly Leu Phe Gly Ser Phe Leu Leu Leu Pro Ser Tyr Leu Gln 450 455 460		
Gln Val Leu His Gln Ser Pro Met Gln Ser Gly Val His Ile Ile Pro 465 470 475 480		
Gln Gly Leu Gly Ala Met Leu Ala Met Pro Ile Ala Gly Ala Met Met 485 490 495		
Asp Arg Arg Gly Pro Ala Lys Ile Val Leu Val Gly Ile Met Leu Ile 500 505 510		
Ala Ala Gly Leu Gly Thr Phe Ala Phe Gly Val Ala Arg Gln Ala Asp 515 520 525		
Tyr Leu Pro Ile Leu Pro Thr Gly Leu Ala Ile Met Gly Met His Met 530 535 540		
Gly Cys Ser Met Met Pro Leu Ser Gly Ala Ala Val Gln Thr Leu Ala 545 550 555 560		
Pro His Gln Ile Ala Arg Gly Ser Thr Leu Ile Ser Val Asn Gln His 565 570 575		
Val Leu Ile Ser Ile Ile Thr Ala Val Met Ser Val Leu Thr Tyr 580 585 590		
Gln Phe Asn His Ser Gln Ile Ile Ala Thr Ala Lys Lys Val Ala Leu 595 600 605		
Thr Pro Gln Ser Gly Ala Gly Arg Gly Ala Ala Val Asn Pro Ser Ser 610 615 620		
Leu Pro Arg Ser Thr Ala Val Thr Thr		

Thr Leu Ile Pro Ala Ala Phe Leu Pro Lys Gln Gln Ala Ser His Arg  
 560 665 670

Arg Ala Pro Leu Leu Ser Ala  
 675

(2) INFORMATION FOR SEQ ID NO:190:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 120 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Thr Pro Glu Lys Ser Phe Val Asp Asp Leu Asp Ile Asp Ser Leu Ser  
 1 5 10 15  
 Met Val Glu Ile Ala Val Gln Thr Glu Asp Lys Tyr Gly Val Lys Ile  
 20 25 30  
 Pro Asp Glu Asp Leu Ala Gly Leu Arg Thr Val Gly Asp Val Val Ala  
 35 40 45  
 Tyr Ile Gln Lys Leu Glu Glu Glu Asn Pro Glu Ala Ala Gln Ala Leu  
 50 55 60  
 Arg Ala Lys Ile Glu Ser Gln Asn Pro Asp Ala Ala Arg Ala Asp Arg  
 65 70 75 80  
 Lys Val Ser Pro Thr Ser Gln Ala Arg Asp Ala Arg Arg Pro Leu Ala  
 85 90 95  
 Arg Ser Ala Arg Leu Ala Tyr Arg Arg Leu Pro Ala Ser Val Pro Thr  
 100 105 110  
 Thr Arg Arg Asp Pro Arg Ile Ala  
 115 120

(2) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:191:

```

                20                25                30
Ile Ala Glu Gly Arg Gln Val Arg Ala Gln Cys Gly Ala Gly Phe Leu
   35                40                45
Glu Arg Arg Pro Ala Val Ser Gly Ala Leu Pro Pro Asn Asn Ala Ser
   50                55                60
Pro Gly Ile Arg Ser Arg Ala Ala Asp Ser Gln Arg His Leu Leu Ala
   65                70                75                80
Gly Asp Gly Ser Asp Val Thr Val Gly
                        85

```

2) INFORMATION FOR SEQ ID NO:192:

- 1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 119 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

x1) SEQUENCE DESCRIPTION: SEQ ID NO:192:

```

Ala Ser Leu Leu Ala Tyr Ser Ala Ala Ala Ala Ser Thr Ala Leu Ala
 1                5                10                15
Val Ala Cys Val Arg Ala Asp His Arg Asp Arg Arg Thr Ile Arg Asp
 20                25                30
His Leu Ala Met Ile His Leu Ala Gln Leu Val Thr Gln Pro Pro Gly
 35                40                45
Ile Val Arg His Arg Leu His His Leu Gly Ile Ala Val Ala Pro Gln
 50                55                60
Pro Gln Gln Val Val Val Leu Ala His His Leu Val Thr Gly Thr Gly
 65                70                75                80
Ala Val Gln His Gln His Arg Val His Val Ala His Val Val Asp Pro
 85                90                95
Leu Asn Gln His Leu Arg Gln Val Leu Arg Pro Ala Arg His Asp Lys
100                105                110
Pro Asp Ala Gly Ile Gly Gln
115

```

3) INFORMATION FOR SEQ ID NO:193:

## X1) SEQUENCE DESCRIPTION: SEQ ID NO:193:

Arg Ala Arg Gly His Arg Ser Ser Lys Gly Ser Arg Trp Ser His Glu  
 1 5 10 15

Val Leu Glu Gly Cys Ile Leu Ala Asp Ser Arg Gln Ser Lys Thr Ala  
 20 25 30

Ala Ser Pro Ser Pro Ser Arg Pro Gln Ser Ser Ser Asn Asn Ser Val  
 35 40 45

Pro Gly Ala Pro Asn Arg Val Ser Phe Ala Lys Leu Arg Glu Pro Leu  
 50 55 60

Glu Val Pro Gly Leu Leu Asp Val Gln Thr Asp Ser Phe Glu Trp Leu  
 65 70 75 80

Ile Gly Ser Pro Arg Trp Arg Glu Ser Ala Ala Glu Arg Gly Asp Val  
 85 90 95

Asn Pro Val Gly Gly Leu Glu Glu Val Leu Tyr Glu Leu Ser Pro Ile  
 100 105 110

Glu Asp Phe Ser  
 115

## 2 INFORMATION FOR SEQ ID NO:194:

1. SEQUENCE CHARACTERISTICS:
  - A. LENGTH: 311 base pairs
  - B. TYPE: nucleic acid
  - C. STRANDEDNESS: single
  - D. TOPOLOGY: linear

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:194

TGCTACGCAG CAATGCTTTT GGTGACAGAT GTGGATGCGG GCGTGGCTGC TGCCGATGGC 50

TGAAAGCGG CGAGCTGTT TCGCGATT TTTGAAAGA TGAAGTGGT TAAAGCGCT 100

TGCGGCGCG CGATGATCG GTTGTGCA GATTCAGCT TCGCGACTG TGAAGAGA 150

TGCTGTGTT GTTGGCTTT GAGGTGCG TCGGCTGCT GGTACCGCG GCGGCGGTT 200

TGATCGGTC GCGGCTGGAT TCGGTGTTA GCGCTGCGG TCGGTACGTC GTGGGCTGG 250

AGCGCTGCT GCGGCGCGG CAGCGGCGA AGCGGTGCT GCGACCGCG TCGGAGCGG 300

TGAGCTGCG TGAAGTAC GCGTGGCT GTTGTGTT TGTGTGAT GTGTGTTA 350

```

TGGTGCTGGC  GTCTCTCATG  GTGGTTTACG  GGCAGGGGCG  CTATGACTGT  CCCCAGCATG      500
GACCCGTCGA  CCGGCTGCGG  CGGCGGGGAG  GCGACCTGGA  CAATGGGGTC  TTCGAGCACC      550
GTTGCCCGGG  GTGCGGCGAG  CCAGTCATCT  GGCAATTGGT  CGACGAAGAT  GCCCCGTTGC      600
GCCCCGCGAG  CCTGTACGCG  GCAGCAAGAC  CGCGCAGGAG  CACTACGCGC  TGGCGTGGTC      700
GGAAACGAAT  GCGCGTTCCG  TGGTGGCGTT  G              811

```

## (2) INFORMATION FOR SEQ ID NO:195:

1. SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 966 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## X1. SEQUENCE DESCRIPTION: SEQ ID NO:195:

```

GTCCCCGAT  GTGGCGGAGT  ATGACTTTCC  GGAACACCGG  CGTAGTAGTC  GAAGATATCG      50
GACTTTGTGG  TCCCGGTGGG  GGGATAGAGC  AGCTGTGGGC  GTTGGTCAGC  GTCACCCGTT      100
GCTCGGACGC  CGAACCCATG  GTTTCACGCT  AGCCTGTGGG  TCACACAAGT  CGCGAGCGTA      150
AGGTGACGGT  GAAATATCGG  GTGGAATTTC  GCGGTGAGGT  TCCGCTCGCG  GACAATCAAG      200
TGATACTCAC  TTACATCCGA  GGCATTTGGA  GGGGTTCGAT  GGTCTTCGGG  CTGGTGAACG      250
TGGCCGTCAA  GTGTACAGC  GTACCGGCG  ACCACGACAT  CAGGTTCCAC  CAGGTGCACG      300
GCAAGGACAA  CGGACCCATG  GGTACAAAGT  GGTGTGTGGA  GCGGTGTGGC  GAGGTGCTCG      350
ATTACCGCGA  TTTTCCCGCG  GGTACGAGT  GCGCGGACGG  GCAAATGGTG  GCGATCACCG      400
AGGACGACAT  GCGCAGCTTG  GGTGAAGAAC  GCAGTCGGGA  GATCGAGGTG  TTGGAGTTCC      450
TCCCCCGCGG  TGAGGTGAG  GCGATGATGT  TCGACCGCAG  GTACTTTTTG  GAGCCTGATT      500
GGAATGCTTG  GAAATGTAAT  GTGTGCTGG  GAAAGACAT  GGTGAGAGAT  GATGAGATCG      550
GATGCTTGG  GTGAGTGAAT  GAAAGAAAT  GAAAGAGTGG  GATGAGATAT      600
TCCCCGTGCA  GTTCCGCTAT  GAGAAAGCGT  AGAAGGATCG  AGACATTGCG  GAGCTGAAGT      650
GAGCGCTGAT  AGAAGCGGCT  GTGCGCGATT  ATGAAACCGA  AAATACCGTT  ACTGATGCCA      700
TGGTGGCTG  TGAGTGAAT  GACCTTTTT  GCGAGGTGG  AGCGCTGGG  TCGCGAGCTG      750
GATGCGGAT  GAAAGTAA  GTTTCGCTG  GAAAGTGGT  GAGGCTGAT  GTATGCTGTT      800

```

1. SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2367 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

CCGCACCGCC GGCAATACCG CCAGCGCCAC CGTTACCGCC GTTTGCGCCG TTGCCCCCGT	50
TCCCGCCCCG CCGCGCGCCG CCGCGGATGG AGTTCTCATC GCCAAAAGTA CTGGCGTTGC	100
CACCGGAGCC GCGGTTGCGG CCGTCACCGC CAGCCCCGCG GACTCCACCG CCCCCACCGA	150
CTCGCGCGCGT GCGACCGTTG CCGCGGTTGC CGATCAACAT GCGGCTGGCG CCACCGTTGC	200
CACCGACCGC AGCGGCTGCG CCGACCGCGC CGACACCAAG CGAGCTGCGG CCGGAGCCAC	250
CATCACCAGC TACCGCACCG ACCGCGCAGA CACCGACGAC CCGGCTGTTG TGAACGTCG	300
CGGTGCGAGT AGCGCGCGCG TTACCGCGAA CCGCACCGCG AACCGCGCGG CCGCGATCGG	350
CGCGCGCGCG GCGGTTGCGG CGGTTGCGCG CGTTGCGGAA CAACAACCGG CCGCGCGCGG	400
CGGTGCGCGG CGCGCGCGCG GTCGCGCGCG CGCGCGCGGAC GCGAAGCGCG CTGCGCGCGT	450
CGCGCGCGATG AGCAGCGTTG CGCGCGACCA CATCGGTTG TCGGTCGCGG TCGGCGGTTG	500
CAACGCTGCG GATCGCGAGG TCGCGCGCGG TCGCGCGCGG CCGCGCGCGT CCGCGCGTAC	550
CACCGATAGC ACCCGCGCGA CCGCGCGCAT CGTTGCGCGG ATCAGCGAAT AGCAACCGCG	600
CGCGCGCGAC ATTGCGCGCG CGTGGCGTGC CGTACGTTG CGCGCGCGAG CCGCGCGTGC	650
CACCGCGCGT ACCACCGAAA CCGCGCGTAC CACCGCGTAA CGTGGCGAGT CGGATGTTTA	700
CGAAGCGCGG CGGTGCGCGG CCGCGCGTAC CACCGCGACT CGCGCGCGGT ACACCGTTG	750
CGCGTTGCG AATCGAGT CGAAGCGG CGGCGATG CGGTGCGG AATCGCGT	800
CGCGCGCGT CCGCGCGCGG CCAATCGCA CGCGCGTAA CGGTGCGG AATCGCGT	850
CGGTGCGGT CCGCGCGCGT CGGTGCGCG CGCGCGGT CGGTGCGGT CGCGCGGT	900
CGGTGCGCG AGTGGCGTGC CCGCGCGTGC CGGTGCGCG CCGCGTTTGA TCACCGATG	950
CGTACGATG TCGCGCGGTG TCGCGCGTGC TCGCGCGCGG CCGCGCGGTG CGATTGACCG	1000
CGGTGCGGT CCGCGCGGTG CCGCGCGGTG TCGCGCGGTG CCGCGCGGTG CCGAAGCGG	1050

```

CGCCGGTACC ACCGGCCCCG CGCTTGCCCG CATTGCCGAT CAACCCGGCG GCGCCTCCGC 1440
TGCCTCCCGT TTGACCGAAC CCGCCAGCCG CGCCGTTGCC ACCGTTGCCA AACAGCAACC 1500
CGCCGGCCCG GCGAGGCTGC CCGGGTGCCG TCCCGTCGGC GCGTTTTCCG ATCAACGGGC 1560
GCCCCAAAAG CGCCTCGGTG GCGGCATTCA CCGCAGCCAG CAGACTCCGC TCAACAGCGG 1620
CTTCAGTGCT GGCATACCGA CCGCGGGCCG CAGTCAACGC CTGCACAAAC TGCTCGTGAA 1680
ACGCTGCCAC CTGTACGCTG AGCGCCTGAT ACTGCCGAGC ATGGGCCCCG AACAAACCCG 1740
CAATCGCCCG CGACACTTCA TCGGCAGCCG CAGCCACCAC TTCCCTCCTC GGGATCGCCG 1800
CGCCCGCATT AGCCCGCCTC ACCTCGCAAC CAATAGTCCA TAAATCCAAA CCGCCAGTTG 1860
CGCCAGCTC CCGCCTCCCG ATCACCAGG ACACCTGCGA CCTCCGGATA CCCCATATCG 1920
CGCCAGCTC TCCCGAGCCG CCACGTGAGC TTCTGCTGCT GCGTGGCCG CCTGACTATC 1980
CGCCCGAGCG CCTCTCTCT GATTCCCGCG CCGCCGCGAG TTGTTGCCCG AGTTGAAGAC 2040
GGGAGGACAG CCGCAGCTTG GTGTAGACGT GGTCAAGTG GGAATGCAG GTCCCGCGCG 2100
AGATGAATAG CCGGACCCCG ATCTCCTCTT TCTGAGTCC CTCACCGAGC AGTAGAGCCA 2160
CCTCAAGCTC TCTCGCTCTC AACCGCCCGC AGCCACTTCT CCGGCTTTT CCGCACCCG 2220
CGCCTGCTG CGGTAGCCG ATCGCTCAT CAGTCGATAA CGCAGTCTCT TCGGCCAGG 2280
CAGCTCGAA ATCGCTCTCA CCGCTCGATT TTGGAAGGCT CGGTAGCGAC CAGTTACAGC 2340
CGCTCTCTA CAGCTCGAA CCGACCG 2357

```

# 1. INFORMATION FOR SEQ ID NO:137

## 1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 175 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## 2. SEQUENCE DESCRIPTION (SEQ ID NO:137)

```

Gln Pro Ala Gly Ala Thr Ile Ala Ala Ser Ser Pro Phe Ala Thr Val
1           5           10           15
Gly Ala Gly Gly Thr Ile Ser Pro Val Thr Thr Gly Thr Ala Ala
2           5           10           15
Thr Thr Gly Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
3           5           10           15

```

Val Ala Thr Ile Thr Ala Lys Gly Ala Arg Asn Val Ala Leu Arg Asp  
65 70 75 80

Ser Ala Val Ala Ala Val Ala Ala Ala Thr Gly Ser Gly Gly Thr  
85 90 95

Ala Val Thr Thr Gly Thr Ala Gly Gly Leu Ala Arg Ala Cys Arg Arg  
100 105 110

Gly Gly Thr Val Ala Ala Gly Ala Thr Gly Arg Arg Ala Gly Ser Ala  
115 120 125

Met Ala Ala Arg Ala Ala Val Ala Ala Gly Leu Ile Thr Asp Ala Gly  
130 135 140

His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly  
145 150 155 160

Ile Asp Pro Val Cys Pro Gly Glu Ala Gly Ala Ala Gly Thr Thr Gly  
165 170 175

Ala Ala Met Ala Glu Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr  
180 185 190

Pro Asp Ala Cys Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala  
195 200 205

Val Ala Pro Gln Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala  
210 215 220

Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Gln Pro Gly Arg  
225 230 235 240

Ala Ser Ala Ala Ala Gly Leu Thr Glu Pro Ala Ser Arg Ala Val Ala  
245 250 255

Thr Val Ala Lys Gln Gln Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys  
260 265 270

His Val Thr Gly Ala Val Ser Asp Gln Arg Ala Thr Gln Lys Arg Leu  
275 280 285

Ala Gly Ala Ser His Arg Thr Thr Gln Thr Pro Leu Ala Ser Gly Asp  
290 295 300

Ser Ala Gly Ile Pro Thr Arg Gly Arg Ser Gln Arg Leu His Lys Leu  
305 310 315 320

Leu Val Ala Arg Val His Leu Thr Ala Thr Arg Leu Ile Leu Pro Ser  
325 330 335

Arg Ser His His Phe Arg Arg Arg Asp Arg Arg Gly Arg Ile Ser Arg  
 355 360 365

Ala His Leu Arg Thr Asn Ser Arg  
 370 375

(2) INFORMATION FOR SEQ ID NO:198:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2852 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:198:

```

33CCAAAACG CCGCGGGGAT CCGCGGCAAC CCGCGGGGCT ACCACCAAGT CTGGGCCCAG      60
GACGTGGCGG CGATGTTTGG CTACCATGCC CGCGCTTCGG CGGCGGTCTC CGCGTTGACA      120
CGGTTCGGCC AGGCGCTGCC GACCGTGGCG CGCGCGGCTC CGCTGCTCAG CGCGGCGCGC      180
GCTCAGGTGA CCACGCGGGT CTTCGCAAC CTGGGCTTGG CGAACGTCCG CGAGGGCAAC      240
GTCCGCAACG GTAATGTCG GAAGTTCAAT CTGGGCTCGG CCAACATCGG CAACGGCAAC      300
ATCGCGAGCG GCAACATCGG CAGGTCCAAC ATCGGTTTTC GCAACGTGGG TCCTGGGTTG      360
ACCGCAGCGG TGAACAACAT CGGTTCGGG AAGACCGGCA GCAACAACAT CGGGTTTGCC      420
ACACCGGCGA GCAACAACAT CGGTTCGGG AATATCGGAG ACUGCAACCG AGGTATCGCG      480
CTACCGGCTA CGGTTCGTT CGGTTCGGG CGGTTCGAAT CGGCGACCGG AACATCGGT      540
CTGTTCAACT CGGCGACCGG AACGTCGGG ATCGGCAACT CGGTATCGG GAATCGGGG      600
ATTGCAACT CGGCGACCGG CTACACACG CGTTTCGCA ACTCGCGCA CGGCAACAG      660
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      720
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      780
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      840
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      900
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      960
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1020
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1080
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1140
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1200
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1260
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1320
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1380
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1440
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1500
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1560
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1620
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1680
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1740
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1800
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1860
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1920
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      1980
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2040
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2100
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2160
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2220
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2280
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2340
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2400
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2460
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2520
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2580
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2640
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2700
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2760
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2820
CTGTTTCGA ACTCGGCAAT AGCGACACG CGGTTCGCA ACTCGCGCA CTACACACG      2852

```

TTGGGTGGGG CACGGGTCTT CAATCTGGGC CTGGCAAAAG GGGGCGTCGT GAACATTCTC 1250  
 GGCAACGCCA ACATCGGCAG TTACAACATT CTCGGGAGCG GAAACCTCGG TACTTTCAAC 1320  
 ATCCTTGGCA GCGGCAACCT GGGCAGGCCA AACATCTTGG GCAGCGGCAG CGTCGGCAGC 1380  
 TTCAATATCG GCAGTGGAAA CATCGGAGTA TTCAATGTGG GTTCGGGAAG CTTGGGAAA 1440  
 TACAACATCG GATCCGGAAA CCTCGGGATC TACAACATCG GTTTTGGAAA CGTCGGCGAC 1500  
 TACAACGTGG GCTTCGGGAA GCGGGGCGAC TTCAACCAAG GCTTTGCCAA CACCGGCAAC 1560  
 AACACATCG GTTTCGCCAA CACCGGCAAC AACACATCG GCATCGGGCT GTCCGGCGAC 1620  
 AACAGCAGG GCTTCAATAT TGCTAGCGGC TGGAACTCGG GCACCGGCAG CAGCGGCTC 1680  
 TTCAATTCGG GCACCAATAA GCTTGGCATC TTCAACGCGG GCACCGGAAA CTCTGGCATC 1740  
 GCAAACTCGG GCACCGGCAG CTGGGTATC GGAACCGCGG GTACCGACAA TACCGGCATC 1800  
 GTCAATCTCG GCAGCTACAA CACGCGCATC GTCAACGCGG GCGACTTCAA CACGGGCTC 1860  
 TACAACGCGG GCAGCTACAA CACCGCGCGC TTCAACGCTG GTACACCAAA CACCGGCGAC 1920  
 TTCAACGTGG GTGACACCA TACCGCGCAG TATAACCGCG GTGACACCA CACCGGCTC 1980  
 TTCAATCTCG GCAACGTCAA TACCGGCGCT TTGACACG GCGACTTCAA CAATGGCTC 2040  
 TTGGTGGGGG GCGATAACCA GCGGCGATC GCGATCGATC TCTCGCTCAG CACTCGATC 2100  
 ATCGGATCAA ATGACGAGAT GGTCAATGAG GTACACAGG TACTGAGCTT GCGCGCAAC 2160  
 ATGATGAGG TACGCGAGG GTGAGCTT TTGCGGCAAA GCTTCTATCT GAGCGGCTC 2220  
 TTCTCTTTCG GCGGGTCAA GCTGAGGCA GCGAGCTCA GCTTCTCAG GATCAGCTC 2280  
 AGGATCGCGG GAGCGAGCTT GAGGCTCGG ATGAGCTT TCGGTCTCTT GAGAGCGCGC 2340  
 GGAATAGCT TGTCAAGAT GATGAGCT GAGGCTT GAGGCTT GAGGCTT 2400  
 GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT 2460  
 GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT 2520  
 GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT 2580  
 GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT 2640  
 GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT GGTGAGCT 2700

GTAAGCGAAT AAACCGAATG GCGGCCTGTC AT

2852

## (2) INFORMATION FOR SEQ ID NO:199:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 943 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

Gly Gln Asn Ala Pro Ala Ile Ala Ala Thr Glu Ala Ala Tyr Asp Gln  
 1 5 10 15  
 Met Trp Ala Gln Asp Val Ala Ala Met Phe Gly Tyr His Ala Gly Ala  
 20 25 30  
 Ser Ala Ala Val Ser Ala Leu Thr Pro Phe Gly Gln Ala Leu Pro Thr  
 35 40 45  
 Val Ala Gly Gly Gly Ala Leu Val Ser Ala Ala Ala Ala Gln Val Thr  
 50 55 60  
 Thr Arg Val Phe Arg Asn Leu Gly Leu Ala Asn Val Arg Glu Gly Asn  
 65 70 75 80  
 Val Arg Asn Gly Asn Val Arg Asn Phe Asn Leu Gly Ser Ala Asn Ile  
 85 90 95  
 Gly Asn Gly Asn Ile Gly Ser Gly Asn Ile Gly Ser Ser Asn Ile Gly  
 100 105 110  
 Phe Gly Asn Val Gly Pro Gly Leu Thr Ala Ala Leu Asn Asn Ile Gly  
 115 120 125  
 Phe Gly Asn Thr Gly Ser Asn Asn Ile Gly Phe Gly Asn Thr Gly Ser  
 130 135 140  
 Asn Asn Ile Gly Phe Gly Asn Thr Gly Asn Ile Gly Ile Gly  
 145 150 155 160  
 Leu Thr Gly Ser Gly Leu Leu Gly Phe Gly Gly Leu Asn Ser Gly Thr  
 165 170 175  
 Ile Asn Ile Gly Leu Phe Asn Ser Gly Thr Gly Asn Val Gly Ile Gly  
 180 185 190  
 Asn Ser Gly Thr Gly Asn Thr Gly Ile Gly Asn Ser Gly Asn Ser Tyr  
 195 200 205

225                      230                      235                      240  
 Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly  
                          245                      250                      255  
 Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu  
                          260                      265                      270  
 Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Ile Thr Gly Asn Phe  
                          275                      280                      285  
 Asn Asn Gly Phe Leu Trp Arg Gly Asp His Gln Gly Leu Ile Phe Gly  
                          290                      295                      300  
 Ser Pro Gly Phe Phe Asn Ser Thr Ser Ala Pro Ser Ser Gly Phe Phe  
                          305                      310                      315                      320  
 Asn Ser Gly Ala Gly Ser Ala Ser Gly Phe Leu Asn Ser Gly Ala Asn  
                          325                      330                      335  
 Asn Ser Gly Phe Phe Asn Ser Ser Ser Gly Ala Ile Gly Asn Ser Gly  
                          340                      345                      350  
 Leu Ala Asn Ala Gly Val Leu Val Ser Gly Val Ile Asn Ser Gly Asn  
                          355                      360                      365  
 Thr Val Ser Gly Leu Phe Asn Met Ser Leu Val Ala Ile Thr Thr Pro  
                          370                      375                      380  
 Ala Leu Ile Ser Gly Phe Phe Asn Thr Gly Ser Asn Met Ser Gly Phe  
                          385                      390                      395                      400  
 Pro Gly Gly Pro Pro Val Phe Asn Leu Gly Leu Ala Asn Arg Gly Val  
                          405                      410                      415  
 Val Asn Ile Leu Gly Asn Ala Asn Ile Gly Asn Tyr Asn Ile Leu Gly  
                          420                      425                      430  
 Ser Gly Asn Val Gly Asn Phe Asn Ile Leu Gly Ser Gly Asn Leu Gly  
                          435                      440                      445  
 Ser Ile Asn Ile Leu Ile Ser Gly Val Val Ile Ser Phe Asn Ile Ile  
                          450                      455                      460  
 Ser Gly Asn Ile Gly Val Phe Asn Val Gly Ser Gly Ser Leu Gly Asn  
                          465                      470                      475                      480  
 Tyr Asn Ile Gly Ser Gly Asn Leu Gly Ile Tyr Asn Ile Gly Phe Gly  
                          485                      490                      495  
 Asn Ile Gly Asn Tyr Asn Val Ile Phe Gly Asn Ala Ile Ser Thr Asn

Gly Asn Asn Asn Ile Gly Ile Gly Leu Ser Gly Asp Asn Gln Gln Gly  
 530 535 540  
 Phe Asn Ile Ala Ser Gly Trp Asn Ser Gly Thr Gly Asn Ser Gly Leu  
 545 550 555 560  
 Phe Asn Ser Gly Thr Asn Asn Val Gly Ile Phe Asn Ala Gly Thr Gly  
 565 570 575  
 Asn Val Gly Ile Ala Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn  
 580 585 590  
 Pro Gly Thr Asp Asn Thr Gly Ile Leu Asn Ala Gly Ser Tyr Asn Thr  
 595 600 605  
 Gly Ile Leu Asn Ala Gly Asp Phe Asn Thr Gly Phe Tyr Asn Thr Gly  
 610 615 620  
 Ser Tyr Asn Thr Gly Gly Phe Asn Val Gly Asn Thr Asn Thr Gly Asn  
 625 630 635 640  
 Phe Asn Val Gly Asp Thr Asn Thr Gly Ser Tyr Asn Pro Gly Asp Thr  
 645 650 655  
 Asn Thr Gly Phe Phe Asn Pro Gly Asn Val Asn Thr Gly Ala Phe Asp  
 660 665 670  
 Thr Gly Asp Phe Asn Asn Gly Phe Leu Val Ala Gly Asp Asn Gln Gly  
 675 680 685  
 His Ile Ala Ile Asp Leu Ser Val Thr Thr Pro Phe Ile Pro Ile Asn  
 690 695 700  
 Gln Gln Met Val Ile Asp Val His Asn Val Met Thr Phe Gly Gly Asn  
 705 710 715 720  
 Met Ile Thr Val Thr His Ala Ser Thr Val Phe Pro Gln Thr Phe Tyr  
 725 730 735  
 Asn Ser Gly Leu Phe Phe Phe Gly Thr Val Asn Leu Ser Ala Ser Thr  
 740 745 750  
 Asn Thr Val Pro Thr Ile Thr Leu Thr Ile Gly Gly Pro Thr Val Thr  
 755 760 765  
 Val Pro Ile Ser Ile Val Gly Ala Ser His Ser Arg Thr Ile Thr Phe  
 770 775 780  
 Leu Gly Ile Asp Ile Ala Pro Gly Ile Gly Asn Ser Thr Thr Asn Pro  
 785 790 795

Asn Val Gly Gly Gly Ser Ser Gly Val Trp Asn Ser Gly Leu Ser Ser  
                     820                    825                    830  
 Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly  
                     835                    840                    845  
 Trp Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr  
                     850                    855                    860  
 Val Asn Leu Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Gly  
                     865                    870                    875                    880  
 Thr Asn Leu Ser Gly Val Phe Arg Gly Pro Thr Gly Thr Ile Phe Asn  
                     885                    890                    895  
 Ala Gly Leu Ala Asn Leu Gly Gln Leu Asn Ile Gly Ser Ala Ser Cys  
                     900                    905                    910  
 Arg Ile Arg His Glu Leu Asp Thr Val Ser Thr Ile Ile Ser Ala Phe  
                     915                    920                    925  
 Cys Gly Ser Ala Ser Asp Glu Ser Asn Pro Gly Ser Val Ser Glu  
                     930                    935                    940

## 2 INFORMATION FOR SEQ ID NO:200:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### (ii) SEQUENCE DESCRIPTION: SEQ ID NO:200

CGATGATAT GGGGATCAT TATGTCATC AGGTGATCGA CATTGATGGG ACC

53

## 3 INFORMATION FOR SEQ ID NO:201:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### (ii) SEQUENCE DESCRIPTION: SEQ ID NO:201

AGTGAATTGA AGGTGAGGT TGGGAGGTT CATTGTAAG GA

42

## 4 INFORMATION FOR SEQ ID NO:202:

### (i) SEQUENCE CHARACTERISTICS:

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:202:

GGATCCTGCA GGCTCGAAAC CACCGAGCGG T

31

## 2) INFORMATION FOR SEQ ID NO:203:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:203:

CTCTGAATTC AGCGCTGGAA ATCGTCGCGA T

31

## 3) INFORMATION FOR SEQ ID NO:204:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:204:

GGATCCAGCG CTGAGATGAA GACCGATCCC GCT

33

## 4) INFORMATION FOR SEQ ID NO:205:

## (1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 38 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:205:

GATATCTTC AGAATTCAGG TTAAAGTTC ATTTCGGA

## 5) INFORMATION FOR SEQ ID NO:206:

## SEQUENCE CHARACTERISTICS

- A LENGTH: 30 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:206:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

CTTCATGGAA TTCTCAGGCC GGTAAGGTCC GCTGCGG

37

## (2) INFORMATION FOR SEQ ID NO:208:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7676 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

TGGCGAATGG GACGCGCCCT GTAGCGCGCG ATTAAGCGCG GCGGGTGTGG TGGTTACCGG 60  
 CAGCGTGACC GGTACACTTG CGAGCGCCCT AGCGCGCCCT CCTTTCGCTT TCTTCGCTTC 120  
 CTTTCTCGCC ACCTTCGCGG GCTTTGCGCG TCAAGCTCTA AATCGGGGGC TCGCTTTAGG 180  
 GTTCGGATTT AGTGGTTTAC CGCAGCTCGA CCGCAAAAAA CTTGATTAGG GTGATGGTTC 240  
 AGCTAGTGGG CGATCGCCCT GATAGAGCGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT 300  
 GTTATATAGT GGACTCTTGT TCGAAATGG GACACACTG AACCTATCTT CGGTCTATTC 360  
 TTTGATTTA TAAGGGATTT TCGCGATTTT CGGTCTATTC TTAATAAATG AGCTGATTTA 420  
 TCGAAATTTT AGCGGAAAT TTAAGAAAT ATTAAGCTTT ACAATTTGAG TTGGGACTTT 480  
 TCGGCGAAAT GTCGCGGAA CCGGTATTTT TTTATTTTTC TAAATACATT GAATATGTA 540  
 GTTCTGATG AGTTAATCTT TAAAGAAAT ATTCAGGAT CAAATGAAAG TCGAATTTAT 600  
 TATATTAAG ATTATGATA TAAATTTT TAAAGAAAG TTTCTGTAT TAAGGAGAA 660  
 TATATTAAG ATCTTCTAT AGAATTTT TAAAGAAAG TTTCTGTAT TAAGGAGAA 720  
 ATTAAGATG AATAGAGCT ATTAATTTT TTTCTGTAT AATAGGTTA TAAAGTGA 780  
 AATAGATG AGTACGAT TAATTTCTT AGAATTTT AAGTTTATG ATTTCTTTTC 840  
 AATATTTT AAGAGCGAA TAAATTTT TTTATTTT AATATTTT TAAAGTGA 900  
 ATTAATTTT AGTATTTT TAAATTTT TAAATTTT TAAATTTT TAAATTTT 960

TGGTGAGTAA CCATGCATCA TCAGGAGTAC GGATAAAATG CTTGATGGTC GGAAGAGGCA 1140  
TAAATTCGGT CAGCCAGTTT AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC 1200  
CTTTGCCATG TTTCAGAAAC AACTCTGGCG CATCGGGCTT CCCATACAAT CGATAGATTG 1260  
TCGCACCTGA TTGCCCCACA TTATCGCGAG CCCATTATA CCCATATAAA TCAGCATCCA 1320  
TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTCCCG TTGAATATGG CTCATAACAC 1380  
CCCTTGATT ACTGTTTATG TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA 1440  
CGTGAGTTTT CTTCCACTG AGCCTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA 1500  
GATCTTTTTT TTCTGCGCGT AATCTGCTGC TTGCAACAA AAAAACCACC GCTACCAGCG 1560  
GTGGTTTGT TTCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC 1620  
AGAGCCGAGA TACCAAATAC TGTCTTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTCAAG 1680  
AATCTCTAG CACCGCTAC ATACCTCGGT CTCTAATCG GTTACCACT GCGTCTGCG 1740  
AGTGGCGATA AGTCTGTCT TACCGGGTGG GACTCAAGAC GATAGTTACC GGATAAGGCG 1800  
CAGCGGTCTG GTTCAACGGG GGTCTCTCTG ACACAGGCGA GCTTGGAGCG AACGACCTAC 1860  
ACCGAAGTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACCTCTCG CGAAGGGAGA 1920  
AAGCGCGACA GGTATCCGGT AAGCGCGAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT 1980  
ACGCGGGGAA ACGCTGCTA TCTTTATAGT CTCTCTCGGT TTGCTCACCT CTGACTTGAG 2040  
ATCGATTTT TCTGATCTC CTGAGCGGCG CGGAGCCTAT CGAAAAACCG CAGCAACCGG 2100  
CCCTTTTAC GCTTCTGCG CTTTCTCTCG CTTTTCTCTG ACATTTCTTT TCTCTCTTA 2160  
TCCCTCATT CTCTGATAA CCGTATTAGC GCTTTGAGT GAGCTGATAC CGCTCGCGCG 2220  
ACCGAAGCA TCGAGCGAG CAGCTAGT AGCGAGAAAG CGGAGAGCG ATCGATCGCG 2280  
GATTTCTCG ATAGGATCT ATAGGATCT TCGAGGAA GATATGCTG AGTCTGATA 2340  
GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT 2400  
GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT 2460  
GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT 2520  
GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT GATCTCTCT 2580

GGTCACTGAT GCCTCCCTGT AAGGGGGATT TGTGTTGATG GGGSTAATGA TACCGATGAA	2750
ACGAGAGAGG ATGCTCACGA TACGGGTAC TGATGATGAA CATGCCCGGT TACTGGAACG	2820
TTGTGAGGGT AAACAACCTG CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG	2880
TCAATGCCAG CGCTTCGTTA ATACAGATGT AGGTGTTCCA CAGGGTAGCC AGCAGCATCC	2940
TGCGATGCAG ATCCGGAACA TAATGGTGCA GGGCGGTGAC TTCCCGCTTT CCAGACTTTA	3000
CGAAACACGG AAACCGAAGA CCATTTCATGT TGTGCTCAG GTCGCAGACG TTTTGCAGCA	3060
GCAGTCGCTT CACGTCGCT CGCGTATCGG TGATTTCATC TGCTAACCG TAAGGCAACC	3120
CCCGCAGCCT AGCCGGGTCC TCAACGACAG GAGCACGATC ATGCGCACCC GTGGGGCCGC	3180
TATCCCGCGG ATAATGCGCT GCTTCTCGCC CAAACCTTTG GTGGCGGGAC CAGTACGAA	3240
GGCTTGAGCG AGGGCGTCCA AGATTCCGAA TACCGCAAGC GACAGGCCCA TCATGCTCGC	3300
GCTCCAGCGA AAGCGGTCTT CCGCGAAAT GACCCAGAGC GGTGCCGGCA CCGTCTCTAC	3360
GATTTGCATG ATAAAGAAGA CAGTCATAAG TCGCGCGAGC ATAGTCATGC CCGCGGCCCA	3420
CGCGAAGGAG CTGACTGGGT TGAAGGCTCT CAAAGGCATC GGTCCGATC CCGGTGCCTA	3480
ATGAGTGAGC TAAGTTACAT TAATTGCTT GCGCTCACTC CCGGCTTTCC AGTCGGGAAA	3540
TTTGTGTGT CAGTTGCATG AATGAATCGG CCAAGCGCCG GGGAGAGGCG GTTTGCGTAT	3600
TGGGCGCCAG GGTGTTTTT TTTTTCACCA CTGAGAGGCG CAACAGCTCA TTGCGCTTCA	3660
TTTGTGCGC CTGAGAGAGT CTGAGCAGC GTTCAAGCT GTTTTGGCGC AGCAGGCCAA	3720
TATGCTGTTT GATGCTGCTT AACGCGCGGA TATTAATGTA GTGCTTTTCC GTATGCTCT	3780
ATCCCACTAC CGAGATATCC ACAGCAAGCG GCAGTCCGGA TTGCGTAATC GCGGCGATTG	3840
ATCCAGCCG CATCTGATCC TTGCGAGCA GATCTCAAT GCGAAGGATG CCGTCAATCA	3900
TAATTCGAT GTTTTGTGA AAACCGAACA TTTAATGTA GTGCGCTTCA ATTTGCGTA	3960
TAATTCGAT GTTTTGTGA AAACCGAACA TTTAATGTA GTGCGCTTCA ATTTGCGTA	4020
AGACAGAACT TAATGGGCGC GCTAACAGCG GATTTGCTG GTGAGCGAAT GCGACGAGAT	4080
GCTGAGCGCG CAGTGGCTTA CGCTTTGAT GCGAGAAAAA AATAGTTTTC ATGGGTGCTT	4140
TTTCAAGAGC ATCAAGAAAT AACGCGCGAA TATTAGTGA GCGAGCTTCC ATAGCAATGC	4200
TAATTCGAT GTTTTGTGA AAACCGAACA TTTAATGTA GTGCGCTTCA ATTTGCGTA	4260

GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCGCC AGTTGTTTGTG 4440  
CCACCCGGTT GGGAAATGTAA TTCAGCTCCG CCATCGCCGC TTCCACTTTT TCCGCGGTTT 4500  
TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC GGTCTGATAA GAGACACCGG 4560  
CATACTCTCC GACATCGTAT AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT 4620  
CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTTGG CCATTTCGATG GTGTCCGGGA 4680  
TCTCGACGCT CTCCTTATG CGACTCCTGC ATTAGGAAGC AGCCCAGTAG TAGGTTGAGG 4740  
CCGTTGAGCA CCGCCCGCCG AAGGAATGCT GCATGCAAGG AGATGGCCGC CAACAGTCCC 4800  
CCGGCCACGG GGCCTGCCAC CATACCCACC CCGAAACAAG CGCTCATGAG CCGGAAGTGG 4860  
TGAGCCCGAT GTTCCCATC TGTGATGTC CCGATATAGG CGCCAGCAAC CGCAGCTGTG 4920  
CGCCCGGTGA TGCCCGCCAC GATGCCCTCC CGGTAGAGGA TCGAGATCTC GATCCCGCA 4980  
AATTAATAGC ACTCACTATA GCGGAATTGT GACCGGATAA CAATTCCTCT CTAGAAATAA 5040  
TTTGTGTTAA GTTTAAGAAG GAGATATACA TATGGGCCAT CATCATCATC ATCACCTGAT 5100  
CGACATCATC GGGACGAGCC CCACATCCTG GGAACAGGCG CGGGCGGAGG CGGTCCAGCC 5160  
CGCCCGGGAT AGGCTGATG ACATCCGCTT CGCTGCGCTC ATTGAGCAGG ACATGCGCTT 5220  
TGACAGCCCC GCGAAGATCA CCTACCGCAT CAAGCTCGAA GTGTCTTCA AGATGAGGCT 5280  
CGCCCAACCC AGGGGCTCGA AAGCACCBAG TGGTTCCCTT AAAACCGGCC CGGGCCCGCC 5340  
TACTGTCCCC ACTACCCGCC CTCTCTGCTT GTGACCTT TCGGAGACCC CTAGCACTCT 5400  
CTCTAAGCC GTTTCGAGC TGTGGGTCTT GCGTTTCTAC GAGAGCTATC CGAAGCTGAC 5460  
GATCAGCCCT CAGGGCACCC GTTCTGGTGC TGGATCCCG CAGGCCGCCG CGGGGACGGT 5520  
AAGATGCGG CCTCCGACC CTATCTCTTC GGAATCTGAT ATGCTGCGCC AAGAGGCTT 5580  
ATGGAATAT CGCTAAGCA CTCTCGTTTA TACCTGAAAT TAAGCTGT ATGGAATGAT 5640  
CGAGGAGCTC AAGCTCAAGC GAAAACTCCT CGCGGATATC TACGAGGACA CTATCAAAAC 5700  
GTGGAAGCA CCGCAGATTC CTGGCTCAA CCGCGCTCTT AAGCTGCGCC GACGCGCTCT 5760  
AGTTGCGCTC CAGCGCTCCG ACGGCTCCCG TGACAGCTTC TTCTTCACCC ACTAGCTCTC 5820  
ATGGAATAT TCGAGCTCTT TCGGAGCTCT TCTGCTCTC GAGAGACCC TGAATCTCT 5880

ACTCGGCGAG GCCCAACTAG GCAATAGCTC TGGCAATTTC TTGTTGCCCC ACSCGCAAAAG 6060  
CATTCAGGCC GCGGCGGCTG GCTTCGCATC GAAAACCCCG GCGAACCAGG CGATTTTCGAT 6120  
GATCGACGGG CCCCCCCCCG ACGGCTACCC GATCATCAAC TACGAGTACG CCATCGTCAA 6180  
CAACCGGCAA AAGGACGCGG CCACCGCGCA GACCTTGCA GCAATTTCTGC ACTGGGCGAT 6240  
CACCAGCGGC AACAAAGGCCT CGTTCTCTGA CCAGGTTTCAT TTCCAGCCCG TGCCGCCCCG 6300  
GGTGGTGAAG TTGTCTGACG CGTTGATCGC GACGATTTCG AGCGCTGAGA TGAAGACCGA 6360  
TSCCCTTACC CTGCGCGCAG AGGCAGGTAA TTTCGAGCGG ATCTCCGGCG ACCTGAAAAC 6420  
CGAGATCGAC CAGGTGGAGT CGACGCGCAG TTCTTTGCA GGCAGTGGC GCGGCGCGGC 6480  
TGGGACGGC GCGGAGGCG CGGTGCTGTC TTTCGAGAA CGAGCGAATA AGCAGAAGCA 6540  
GGAATCTGAC GAGATCTGCA CGAATATTCC TCAGGCGCGC GTGCAATACT CGAGGGCGGA 6600  
CGAGAGCGAG CAGCAGGCGC TGCTCTCGCA AATGGGCTTT GTGCGCACAA CGGCGCGCTC 6660  
TCCGCGCTCG ACCGCTGCAG CGCCACCGCG ACCGCGGACA CCGTTTGCGC CCGCACACC 6720  
GGCGCGCGCG AACACGCGCA ATGCGCGAGC GGGCGATCCC AACGCGAGCAG CTCGCGCGGC 6780  
CGACCGGAAC GCACCGCGCG CAGCTGTGAT TCGCGCAAAAC GCACCGCAAC CTGTCCGGAT 6840  
CGACAACCGG GTTGGAGGAT TCAGCTTCCG CCGCGCTGCT GGCTGGGTGG AGTCTGACCG 6900  
CGCGCACTTC GACTAGCGTT CAGTACTCTT CAGCAAAACG ACCGCGGAGC CGGCATTTCC 6960  
CGGACAGCGG CCGCGCGTCC CGAATGAGAG CGGTATCGTC CTGCGCGCGC TAGACCAAAA 7020  
CGTTTACCGC ACCCGCGAAG CGAGCGATTC CAAGCGCGCG CCGCGCTTCC CCGCGAGAT 7080  
CGGTGACTTC TATATCGCGT ACCCGCGGAC CGCGATCAAC CAGGAAACCG TCTCGCTTGA 7140  
TGGCAACCGG GTGTCTCGAA CGCGCTGCTA TTACGAGCTC AAGTTGAGCG ATCGAGTAA 7200  
CGGCAACCGG GAGATCTGCA CGGATCTGAT CGGTCTGCGC CGGCGAAGC CAGCGGACCG 7260  
CGGCAACCGG TACGCTGAT TTCTGATGTC CGTGGGAGT CGGCAAAACG CGGTGAGCAA 7320  
CGGCGCGCGG AAGCGCGTCC CGGATCTGAT CGGTCTGCTG CTGCGCGCGC CGGTGCGCGC 7380  
CGGCGCGCGT CTGCGAGAGC CGGTCTGCGC CGGCGCGCGC CGGCGCGAAG TCGTCTCTAC 7440  
CGGCAACGAA CGGATCTGCA AAGCGAGCTT AAGCGCTGCA AAGTTCTGCA GATCTGATC 7500  
CGGCAACGAA CGGTCTGAT AAGCGAGCTT AAGCGCTGCA AAGTTCTGCA GATCTGATC 7560

## (2) INFORMATION FOR SEQ ID NO:209:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 802 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

```

Met Gly His His His His His Val Ile Asp Ile Ile Gly Thr Ser
 1             5             10             15

Pro Thr Ser Trp Glu Gln Ala Ala Ala Glu Ala Val Gln Arg Ala Arg
 20             25             30

Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
 35             40             45

Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
 50             55             60

Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
 65             70             75             80

Gly Ser Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro
 85             90             95

Ala Ser Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr
100            105            110

Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn
115            120            125

Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln
130            135            140

Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Thr Leu Ser
145            150            155

Glu Gly Asn Met Ala Ala His Lys Glu Leu Met Asn Leu Ala Leu Ala
160            165            170            175

Ile Ser Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His
180            185            190

Leu Lys Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile
195            200            205

```

Asp Thr Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly  
 245 250 255  
 Trp Gly Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val  
 260 265 270  
 Pro Gly Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys  
 275 280 285  
 Ala Glu Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp  
 290 295 300  
 Gln Ala Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser  
 305 310 315 320  
 Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Glu Ala Ala Ala Ala  
 325 330 335  
 Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp  
 340 345 350  
 Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile  
 355 360 365  
 Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala  
 370 375 380  
 Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp  
 385 390 395 400  
 Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp  
 405 410 415  
 Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala  
 420 425 430  
 Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile Ser Gly Asp Leu  
 435 440 445  
 Lys Thr Gln Val Asp Gln Val Gln Ser Thr Ala Gly Ser Leu His Gln  
 450 455 460  
 Gln Ser Arg Gln Ala Ala Ser Ser Ala Ala Gln Ala Ala Val Val Arg  
 465 470 475 480  
 Phe Gln Gln Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser  
 485 490 495  
 Thr Asp Gln Arg His Ala Gly Thr Thr Tyr Ser Arg Ala Asp Glu Gln  
 500 505 510

Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro  
 530 535 540

Val Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro  
 545 550 555 560

Gly Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro  
 565 570 575

Pro Pro Val Ile Ala Pro Asn Ala Pro Gln Pro Val Arg Ile Asp Asn  
 580 585 590

Pro Val Gly Gly Phe Ser Phe Ala Leu Pro Ala Gly Trp Val Glu Ser  
 595 600 605

Asp Ala Ala His Phe Asp Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr  
 610 615 620

Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr  
 625 630 635 640

Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu  
 645 650 655

Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu  
 660 665 670

Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser  
 675 680 685

Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys  
 690 695 700

Phe Ser Asp Pro Ser Lys Pro Asn Gly Ile Ile Trp Thr Gly Val Ile  
 705 710 715 720

Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp  
 725 730 735

Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asn Lys Glu Ala  
 740 745 750

Val Val Ala Leu Ala Gln Ser Ala Asn Pro Leu Val Ala Pro Pro Pro  
 755 760 765

Ala Pro Ala Pro Ala Pro Ala Glu Trp Ala Pro Ala Pro Ala Pro Ala  
 770 775 780

Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu  
 785 790 795 800

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 454 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

```

GTGGCGGCGC TGCGGCGCGC CAGCAGAGCG ATGTGCATCC GTTCGCGAAC CTGATCGCGG    60
TCGACGATGA GCGCGCCGAA CGCCGCGACG ACGAAGAACG TCAGGAAGCC GTCCAGCAGC    120
GCGGTCCGCG CGGTGACGAA GCTGACCCCG TCGCAGATCA GCAGCACCCC GCGGATGGCG    180
CCGACCAATG TCGACCGGCT GATCCGCGCG ACGATCCGCA CCACCAGCGC CACCAGGACC    240
ACACCCAGCA GGGCGCGGCT GAACCGCCAG CCGAATCCGT TGTGACCGAA GATGGCCTCC    300
TCGATCCGCA TCAGCTGCTT ACCGACCGCG GGTGAACCA CCAGGCCGTA CCCGGGGTTG    360
TCTTCCACCC CATGGTTGTT CAGCACCTGC CAGGCCTGGC GGTGCGTAAT GCTTCTCCTC    420
GAAGATGGGG GTGCCGGCAT CCGTCACCGA GCCC                                     454

```

## (x) INFORMATION FOR SEQ ID NO:211:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

```

TCCAGAAGTA TCGCGGATCC TCGGTGGCGG ACGCGGAACG GATTCGCGCG GTCCCGGAAC    60
GCATCGTCCG CACCAAGAAG CAAGGCAATG ACCTCGTCTG TGTCTCTCTT GCCATGGGGG    120
ATACCAACCGA CGACCTGCTG GATGTGGCTG AGCAGGTGTG CCGCGCGCGG CGGCTCTGGG    180
ACCTCGACAT TCTGCTTACG GCCGGTGAAC GCATCTCGAA TGCCTTGGTG GCCATGGGCA    240
TCGAGTCCGT TCGCGCGCAT GCGCGGTCTG TCACC3GTTT GCAGGCCCGG GTGATCACCA    300
TCCGCAACCA CCGCAACCGG AAGATCATCG ACCTCACGCG GGGGCGGCTG CAAACCGCGC    360
TTGAGGAAGG CCGGCTCTCT TTGGTGGCGG GATTCGAAGG GGTGAGCCAG GACACCAAGG    420
ATGTCAACGAC GTTGGGCGCG GCGGGCTCGG ACACCACCGG GTCGCGCATG                                     470

```

## (x) INFORMATION FOR SEQ ID NO:212:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 272 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

CCCAGCTCCT CAAGGACGCG GAGAGCGATG AAGTCTTGGG CAAAATGAAG GTGTCTGCGC 240  
TGGTTGAGGC GTTGCCAAAG GTGGGCAAGG TCCAGGGCG 279

(2) INFORMATION FOR SEQ ID NO:213:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:213:

ACACGGTCCA ACTCGACGAG GGCCTCGTGG AGGTGTCGAC CGACAAGGTC GACACCGAAA 60  
TCCCTCGCCG GCGGCGGGTG TGCTGAUCAA GATCATCGCC CAAGAAGATG ACACGGTCCA 120  
AGTCGGCGGC GAGGTCTCTG TCATTGGGGA CGCCCATGAT GCGGCGCAGG CCGCGGTCCC 180  
GGACCCGAG AAAGTCTCTG CCGGCCCAAC TCGAATCCA 219

(2) INFORMATION FOR SEQ ID NO:214:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 342 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:214:

TGCTGCGCGA CATCGGCGCC GCGCCCGCCG TGAAGCGGCG ACCTAAGGCC GTCCCGGAGC 60  
GAGCCCGGAC GCGGAAGGCC GAACCCGCAJ CATCGCGGCG GCGGCGCGAG CGAGCCGCTG 120  
TGGCCGAGGG GGCACCTTAC GTGACCGCCG TGCTGCGAAA GCTGGCGCTG GAAAACAACA 180  
TGGACCTGCG CCGGCTGACC GGCACCGGAG TGCTGCTGCG CATCGCGCAA TAGGATGTGG 240  
TGGCCCGCGC TGAACAAAAG AAGCGGGCGA AAGCACCGCG GCGGCGCGCG TAGGCCGCTG 300  
TGGCCCGCGC CCGGAAAGCG CCGCTGAAG ATCGGATGCG GC 342

(2) INFORMATION FOR SEQ ID NO:215:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:215:

CGGAAAACAA CATCGACCTC GCGGGGGTGA CCGGCACCGG AGTGGGTGGT CGCATCCGCA	300
AACAGGATGT GCTGGCCGCG GCTGAACAAA AGAAGCGGGC GAAAGCACCG GCGCCCTGAG	360
CGCTTCATCA CCGCGTTAAC CAGCTTGCCC CAGAAGCCCG CTTCGACCTC TTCGCGGGTC	420
TTGGTCCGCT GCAGGCGGTC GCGGAGCCAG TTCAGGTTAG GCGGCCGAAA TCTTCCAGTT	480
CGCCAGGAAG GGCACCCGGA ACAGGGTCCG CACCC	515

## (2) INFORMATION FOR SEQ ID NO:216:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 557 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## xii SEQUENCE DESCRIPTION: SEQ ID NO:216:

TGGACCCCAA BGTGCAGATT CACAGGCCA TTGAGGAAGC ACAGCGCACC CACCAAGCTC	60
TGACTCAACA GCGCGCCCAA GTGATCGGTA ACCAGCGTCA ATTGGAGATG CGACTCAAGC	120
GACAGCTGGC GGATATCGAA AAGCTTCAGC TCAATGTGCG CCAAGCCCTG ACGCTGGCCG	180
ACCAGGCCAC GCGCGCCGGA GACGCTGCCA AGGCCACCGA ATACAACAAC GCGCGCGAGG	240
CGTTGCGAGC CCAGCTGGTG ACCGCGGAGC AGAGCGTCGA AGACCTCAAG ACGCTGCATG	300
ACCAGCGCGT TAGCGCCGCA GCTCAGGCCA AGAAGCCCGT CGAACGAAAT GCGATGGTGC	360
TGCAGCAGAA GATCGCCGAG CGAACCAAGC TCTCAGGCCA GCTCGAGCAG GCGAAGATGC	420
AGGAGCAGGT CAGCGCATCG TTGCGGTGCA TCACTGAGCT CGCCGCGCCA GGCAACACGC	480
CGAGCCTCGA CGAGGTGCGC GACAAGATCG AGCCTCGCTA CGCCAACGCG ATCGGTTCCG	540
CTGAACCTGC CGAGAGT	557

## (2) INFORMATION FOR SEQ ID NO:217:

## 1 SEQUENCE CHARACTERISTICS:

- A LENGTH: 223 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

## 11 MOLECULE TYPE: Genomic DNA

## x1 SEQUENCE DESCRIPTION: SEQ ID NO:217:

TAGATAGGT TTGCAATCC AGCTGGTTT TTAAGGAT TGGGACCTT GTGATAGGCT	60
TAAGTGGAC TTCCCGGAG CAGGATGCA TTAAAGATT AACAGAAATG GCTTCTGCG	120
TAAATGCG GCGACTCGGT CAGAGGCTCA CTAAGGGAG GATTAGGCG TGGTCAAAAC	180
AAGAAAGCGA CACCGTCCAA CTCGAGGAG CCGTCTGGA GTT	223

## 2 INFORMATION FOR SEQ ID NO:218:

## 1 SEQUENCE CHARACTERISTICS:

- A LENGTH: 174 base pairs
- B TYPE: nucleic acid

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:218:

AAGAAGTACA	TCTGCCGGTC	GATGTCCGGC	AACCACGGCA	GCCAACCGGC	GCAGTAGCCG	60
ACCAGGACCA	CCGCATAACG	CCAGTCCCGG	CGCACAAACA	TACGCCACCC	CGCGTATGCC	120
AGGACTGGCA	CCGCCAGCCA	CCACATCGCG	GGCGTGCCGA	CCAGCATCTC	GGCCTTGACG	180
CACCACTGTG	CGCCGCAGCC	TGCAACGTCT	TGCTGGTCTG	TGGCGTACAG	CACCGGCCCG	240
AACGACATGG	GCCAGGTCCA	CGGTTTGGAT	TCCCAAGGGT	GGTAGTTGCC	TGCGGAATTC	300
GTCCAGGCCCG	CGTGAAGTGC	GAACGCTTTG	GCGGTGTATT	GCCAGAGCGA	GCGCACGGCG	360
TCGGGCAGCG	GAACAACCGA	GTTGCGACCG	ACCGCTTGAC	CGACCGCATG	CCGATCGATC	420
GCGGTCTCGG	ACGCGAACCA	CGGAGCGTAG	GTGCCCAGAT	AGACCGCGAA	CGGGATCAAC	480
CCCAGCGCAT	ACCCGCTGGG	AAGCACGTCA	CGCCGCACTG	TTCCCAGCCA	CGGTCTTTTC	540
ACTTGGTATG	AACGTCCGGC	CGCCACGTCA	ACGCCCAGC			578

## (2) INFORMATION FOR SEQ ID NO:219:

## 1. SEQUENCE CHARACTERISTICS:

- A) LENGTH: 484 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

## (11) MOLECULE TYPE: Genomic DNA

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:219:

ACAACGATCG	ATTGATATCG	ATGAGAGACG	GAGGAATCGT	GGCCCTTCCC	CAGTTGACCG	60
ACGAGCAGCG	CGCGGCCGCG	TTGGAGAAGG	CTGCTGCCCG	ACGTCCGAGC	CGAGCAGAGC	120
TCAAGGATCG	GCTCAAGCGT	GGCGGCACCA	ACCTCACCCA	GCTCCTCAAG	GACGCGGAGA	180
GCGATGAAGT	CTTGGGCAAA	ATGAAGGTGT	CTGCCCTGCT	TGAGGCTTTC	CCAAAGGTGG	240
GCAAGGTCAA	GGCGCAGGAG	ATCATGACCG	AGCTGGAAAT	TGCGCCCGAG	CCCGCCGCGT	300
TGGTGGCCTC	GGTGACCGTC	AGCCCAAGGC	TGTGCTGGAA	AAGTTCCGCT	CCGCGTAACC	360
CCCGCCGCGG	ACGATGCGCG	CCGGAAGGCG	TGTGCTGGCG	GTACCCCGCG	ATACGGGGGA	420
TAAGCGCGCT	GACAGGGCCA	GTCACAAATT	TAGGCGGAA	CCCGGCTGG	GCGGGAACCC	480
CCCC						484

## 2. INFORMATION FOR SEQ ID NO:220:

## 1. SEQUENCE CHARACTERISTICS:

- A) LENGTH: 537 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

## (11) MOLECULE TYPE: Genomic DNA

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:220:

AGGACTGACA	CGCCGAGCCA	CCACATCCCG	GGCGTGCCGA	CCAGCATCTC	GGCCTTGACG	60
AGGAAATTC	CCCGGAGGCC	TCCAACTCTT	TGCTGGTCTG	TGGCGTACAG	CACCGGCCCG	120
AACGACATGG	CCGAGGTCCA	CGGTTTGGAT	TCCCAAGGGT	GGTAGTTGCC	TGCGGAATTC	180

ACCTGCTACT GACGTCGCGC CGCCACGTCG AACGCCAGCG CCATCGCGCC GAAGAACAGC 480  
ACGAASTACA CGCCGGACCA CTGGGTGGCG CAAGCCAATC CCAAGCAGCA CCCCAGC 537

## (2) INFORMATION FOR SEQ ID NO:221:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

Gly Gly Ala Ala Ala Gly Gln Gln Ser Asp Val His Pro Phe Ala Asn  
1 5 10 15  
Leu Ile Ala Val Asp Asp Glu Arg Ala Glu Arg Arg Asp Asp Glu Glu  
20 25 30  
Arg Gln Glu Ala Val Gln Gln Arg Gly Pro Arg Gly Asp Glu Ala Asp  
35 40 45  
Pro Val Ala Asp Gln Gln His Pro Gly Asp Gly Ala Asp Gln Cys Arg  
50 55 60  
Pro Ala Asp Pro Pro His Asp Pro His His Gln Arg His Gln Asp His  
65 70 75 80  
Thr Gln Gln Gly Ala Gly Glu Pro Pro Ala Glu Ser Val Val Thr Glu  
85 90 95  
Asp Gly Leu Pro Asp Arg Asp Gln Leu Leu Thr Asp Arg Arg Val Asn  
100 105 110  
His Gln Ala Val Pro Gly Val Val Phe His Pro Met Val Val Gln His  
115 120 125  
Leu Pro Gly Leu Ala Val Arg  
130 135

## (2) INFORMATION FOR SEQ ID NO:222:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 156 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

His Lys Tyr Gly Gly Ser Ser Val Ala Asp Ala Glu Arg Ile Arg Arg  
1 5 10 15  
Val Ala Glu Arg Ile Val Ala Thr Lys Lys Gln Gly Asn Asp Val Val  
20 25 30  
Val Thr Val Ser Ala Met Glu Asp Thr Thr Asp Asp Leu Leu Asp Leu  
35 40 45 50 55 60

```

65              70              75              80
Gln Ser Leu Gly Ala His Ala Arg Ser Phe Thr Gly Ser Gln Ala Gly
              85              90              95
Val Ile Thr Thr Gly Thr His Gly Asn Ala Lys Ile Ile Asp Val Thr
              100              105              110
Pro Gly Arg Leu Gln Thr Ala Leu Glu Glu Gly Arg Val Val Leu Val
              115              120              125
Ala Gly Phe Gln Gly Val Ser Gln Asp Thr Lys Asp Val Thr Thr Leu
              130              135              140
Gly Arg Gly Gly Ser Asp Thr Thr Ala Val Ala Met
145              150              155

```

(2) INFORMATION FOR SEQ ID NO:223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

```

Pro Ala Tyr Pro Ala Gly Thr Asn Asn Asp Arg Leu Ile Ser Met Arg
 1           5           10           15
Asp Gly Gly Ile Val Ala Leu Pro Gln Leu Thr Asp Glu Gln Arg Ala
          20           25           30
Ala Ala Leu Glu Lys Ala Ala Ala Arg Arg Ala Arg Ala Glu Leu
          35           40           45
Lys Asp Arg Leu Lys Arg Gly Gly Thr Asn Leu Thr Gln Val Leu Lys
          50           55           60
Asn Ala Glu Ser Asp Gln Val Leu Gly Lys Met Lys Val Ser Ala Leu
          65           70           75           80
Leu Glu Ala Leu Pro Lys Val Gly Lys Val Gln Ala
          85           90

```

(2) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

```

Thr Val Glu Leu Asp Glu Pro Leu Val Gln Val Leu Thr Asp Lys Val
          1           5           10

```

Gly Asp Ala His Asp Ala Gly Glu Ala Ala Val Pro Ala Pro Gln Lys  
 50 55 60  
 Val Ser Ala Gly Pro Thr Arg Ile  
 65 70

## (2) INFORMATION FOR SEQ ID NO:225:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 113 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

Ala Ala Asp Ile Gly Ala Ala Pro Ala Pro Lys Pro Ala Pro Lys Pro  
 1 5 10 15  
 Val Pro Glu Pro Ala Pro Thr Pro Lys Ala Glu Pro Ala Pro Ser Pro  
 20 25 30  
 Pro Ala Ala Gln Pro Ala Gly Ala Ala Glu Gly Ala Pro Tyr Val Thr  
 35 40 45  
 Pro Leu Val Arg Lys Leu Ala Ser Glu Asn Asn Ile Asp Leu Ala Gly  
 50 55 60  
 Val Thr Gly Thr Gly Val Gly Gly Arg Ile Arg Lys Gln Asp Val Leu  
 65 70 75 80  
 Ala Ala Ala Glu Gln Lys Lys Arg Ala Lys Ala Pro Ala Pro Ala Ala  
 85 90 95  
 Gln Ala Ala Ala Pro Ala Pro Lys Ala Pro Pro Glu Asp Pro Met  
 100 105 110  
 Pro

## 2. INFORMATION FOR SEQ ID NO:226:

## i. SEQUENCE CHARACTERISTICS:

- A) LENGTH: 118 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

## ii. MOLECULE TYPE: protein

## xi. SEQUENCE DESCRIPTION: SEQ ID NO:226:

Ala Leu Val Ser Ile Ser Ala Asp Glu Asp Ala Thr Val Pro Val Gly  
 1 5 10 15  
 Gly Glu Leu Ala Arg Ile Gly Val Ala Ala Asp Ile Gly Ala Ala Pro  
 20 25 30  
 Val Thr Pro Ala Pro Leu Pro Val Pro Glu Pro Ala Pro Thr Pro  
 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 118

```

65           70           75           80
Glu Asn Asn Ile Asp Leu Ala Gly Val Thr Gly Thr Gly Val Gly Gly
           85           90           95
Arg Ile Arg Lys Gln Asp Val Leu Ala Ala Ala Glu Gln Lys Lys Arg
           100           105           110
Ala Lys Ala Pro Ala Pro
           115

```

## (2) INFORMATION FOR SEQ ID NO:227:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 185 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (iii) SEQUENCE DESCRIPTION: SEQ ID NO:227:

```

Asp Pro Lys Val Gln Ile Gln Gln Ala Ile Glu Glu Ala Gln Arg Thr
1           5           10           15
His Gln Ala Leu Thr Gln Gln Ala Ala Gln Val Ile Gly Asn Gln Arg
20           25           30
Gln Leu Glu Met Arg Leu Asn Arg Gln Leu Ala Asp Ile Glu Lys Leu
35           40           45
Gln Val Asn Val Arg Gln Ala Leu Thr Leu Ala Asp Gln Ala Thr Ala
50           55           60
Ala Gly Asp Ala Ala Lys Ala Thr Glu Tyr Asn Asn Ala Ala Glu Ala
65           70           75           80
Phe Ala Ala Gln Leu Val Thr Ala Glu Gln Ser Val Glu Asp Leu Lys
85           90           95
Thr Leu His Asp Gln Ala Leu Ser Ala Ala Ala Gln Ala Lys Lys Ala
100           105           110
Val Gln Arg Asn Ala Met Val Leu Gln Gln Lys Ile Ala Glu Arg Thr
115           120           125
Lys Leu Leu Ser Gln Leu Glu Gln Ala Lys Met Gln Glu Gln Val Ser
130           135           140
Ala Ser Leu Arg Ser Met Ser Glu Leu Ala Ala Pro Gly Asn Thr Pro
145           150           155           160
Ser Leu Asp Glu Val Arg Asp Lys Ile Glu Arg Arg Tyr Ala Asn Ala
165           170           175
Glu Ala Glu Leu Ala Glu Ser
180           185

```

## (2) INFORMATION FOR SEQ ID NO:228:

## (i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 71 amino acids
- B TYPE: amino acid
- C STRANDEDNESS: single

## (x1) SEQUENCE DESCRIPTION SEQ ID NO:228:

```

Val Ser Thr Ser Thr Trp Val Pro His Pro Val Arg Asp Arg Val Ile
 1             5             10             15
Gly Gln Arg Trp Thr Cys Ala Asp Arg Arg Ser Ile Glu Glu Ser Thr
             20             25             30
Glu Met Ala Phe Ser Val Gln Met Pro Ala Leu Gly Glu Ser Val Thr
             35             40             45
Glu Gly Thr Val Thr Arg Trp Leu Lys Gln Glu Gly Asp Thr Val Glu
             50             55             60
Leu Asp Glu Pro Leu Val Glu
             65             70

```

## (2) INFORMATION FOR SEQ ID NO:229:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (11) MOLECULE TYPE: protein

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:229:

```

Glu Val His Leu Pro Val Asp Val Gly Glu Pro Arg Gln Pro Thr Gly
 1             5             10             15
Ala Val Ala Asp Gln Asp His Arg Ile Thr Pro Val Pro Ala His Lys
             20             25             30
His Thr Pro Pro Arg Val Cys Gln Asp Trp His Arg Gln Pro Pro His
             35             40             45
Arg Gly Arg Ala Asp Gln His Leu Gly Leu Asp Ala Arg Leu Cys Ala
             50             55             60
Ala Ala Cys Asn Val Leu Leu Val Asp Gly Val Gln His Arg Pro Gln
             65             70             75             80
Arg His Gly Pro Gly Pro Arg Phe Gly Phe Pro Arg Val Val Val Ala
             85             90             95             100
Cys Gly Ile Arg Gln Ala Arg Val Glu Val Gln Arg Phe Gly Gly Val
             105             110             115
Leu Pro His Arg Ala Val Thr Val Gly His Arg Asn Asn Arg Val Ala
             120             125             130
Thr Asp Arg Leu Thr Ser Arg Met Pro Leu Asp Arg Gly Leu His Arg
             135             140             145
Gln Pro Arg Ser Val His Gly Gln Ile Asp Arg Gln Arg Asp Gln Pro
             150             155             160
Gln Arg Ile Pro Ala Gly Lys His Val Thr Pro His Cys Ser Gln Pro
             165             170             175
Arg Ser Leu His Leu Val
             180

```

## (2) INFORMATION FOR SEQ ID NO:230:

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(12) SEQUENCE DESCRIPTION: SEQ ID NO:230:

```

Asn Asp Arg Leu Ile Ser Met Arg Asp Gly Gly Ile Val Ala Leu Pro
 1           5           10           15
Gln Leu Thr Asp Glu Gln Arg Ala Ala Leu Glu Lys Ala Ala Ala
 20           25           30
Ala Arg Arg Ala Arg Ala Glu Leu Lys Asp Arg Leu Lys Arg Gly Gly
 35           40           45
Thr Asn Leu Thr Gln Val Leu Lys Asp Ala Glu Ser Asp Glu Val Leu
 50           55           60
Gly Lys Met Lys Val Ser Ala Leu Leu Glu Ala Leu Pro Lys Val Gly
 65           70           75           80
Lys Val Lys Ala Gln Glu Ile Met Thr Glu Leu Glu Ile Ala Pro His
 85           90           95
Pro Ala Ala Phe Val Ala Ser Val Thr Val Ser Ala Arg Pro Cys Trp
100          105          110
Lys Ser Ser Ala Pro Pro Asn Pro Ala Gly Arg Arg Cys Gly Pro Glu
115          120          125
Gly Leu Trp Trp Ala Tyr Pro Arg Ile Arg Gly Arg Ser Gly Leu Thr
130          135          140
Gly Pro Ala His Asn Ser Gly Arg Thr Pro Arg Trp Gly Gly Thr Arg
145          150          155          160

```

(13) INFORMATION FOR SEQ ID NO:231:

1. SEQUENCE CHARACTERISTICS:

A. LENGTH: 178 amino acids

B. TYPE: amino acid

C. STRANDEDNESS: single

D. TOPOLOGY: linear

11. MOLECULE TYPE: protein

12. SEQUENCE DESCRIPTION: SEQ ID NO:231:

```

Asn Ile His Arg Ile Pro Pro His Ala Leu Arg Ala Arg Gln His Leu
 1           5           10           15
Ile Leu Asp Ala Arg Leu Pro Ala Ala Ala Thr Asn Val Leu Leu Val
 20           25           30
Asp Gly Val Gln His Arg Pro Gln Arg His Gly Pro Gly Pro Arg Phe
 35           40           45
Ily Phe Pro Arg Val Val Val Ala Cys Gly Ile Arg Gln Ala Arg Val
 50           55           60
Ile Val Glu Arg Pro Gly Gly Val Val Pro Glu Arg Ala His Gln Val
 65           70           75           80

```

```

      115              120              125
Val Thr Pro His Cys Pro Gln Pro Arg Ser Leu His Leu Val Leu Thr
      130              135              140
Ser Arg Arg His Val Glu Arg Gln Arg His Arg Ala Glu Glu Gln His
      145              150              155              160
Glu Val His Ala Gly Pro Leu Gly Gly Ala Ser Gln Ser Gln Ala Ala
      165              170              175
Pro Arg

```

## (2) INFORMATION FOR SEQ ID NO:232:

## 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 11. MOLECULE TYPE: Genomic DNA

## 1X1. SEQUENCE DESCRIPTION: SEQ ID NO:232:

```

ATGCCAAGCC GGTGCTGATS CCGGAGCTCC CGCAATCGGT GACCGAGGGG ACCGTCATTC      60
GTTGGCTGAA GAAGATCGGG GATTCGGTTC AGGTTGACGA GCCACTCGTG GAGGTGTCCA      120
CCGACAAGGT GGACACCGAG ATCCCGTCCC CGGTGGGTGG GGTCTTGSTC AGTATCAGCC      180
CCGACGAGGA CCGCACGGTG CCCGTCGGCC CGAGTTGGC CCGGATCGGT CTCGCTGCCC      240
AGATCGGCGC CCGCGCCGCG CCGAAGCCCT C

```

## (2) INFORMATION FOR SEQ ID NO:233:

## 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 11. MOLECULE TYPE: protein

## 1X1. SEQUENCE DESCRIPTION: SEQ ID NO:233:

```

Met Asp Phe Val Leu Met Phe Glu Asp Ile His Ser Val Thr Gln Val
10
Glu Val Leu Asp Tyr Leu Leu Leu Ile Glu Asp Ser Val His Val Asp
20
Phe Pro Leu Val Glu Val Ser Thr Asp Lys Val Asp Thr Glu Ile Pro
30
Ser Pro Val Ala Gly Val Leu Val Ser Ile Ser Ala Asp Glu Asp Ala
40
Thr Val Pro Val Gly His Glu Leu Ala Arg Ile Thr Val Ala Ala Glu
50

```

(A) LENGTH: 107 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

(2) INFORMATION FOR SEQ ID NO:235:

(A) LENGTH: 330 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION. SEQ ID NO:235.

2 INFORMATION FOR SEQ ID NO: 216

A) LENGTH: 111 amino acids  
B) TYPE: amino acid  
C) STRANDEDNESS: single  
D) TOPOLOGY: linear

A1 SEQUENCE DISPOSITION: 300 11 40 210

Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala  
1 5 10 15  
Ala Ala Leu Val Phe Pro Ala Ser Val Ala Ser Ala Asp Pro Phe Asp  
20 25 30  
Pro His Glu Pro Asp Met Thr Lys Gly Thr Lys Pro Leu Gly Arg Trp  
35 40 45 50  
Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala  
55 60 65 70 75 80 85 90 95 100

35 90 95  
 Pro Gly Gly Cys Gly Gly Ala Ile Pro Ser Glu Gln Pro Asn Ala Pro  
 100 105 110

## 2. INFORMATION FOR SEQ ID NO:237:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: cDNA

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

GTGACCAAGG	TGGGCGCTGG	ACCAACCCGG	GCAGCCGCGAG	CGGCGGCGGG	TCCGCGCGGT	60
CCGCGCGCAA	CGGTGGCGCG	GGGGGTAACG	CCACCGGCTC	AGGCGGCAAG	GGCGCGCGCG	120
GTGGCAATCG	CGGTGATCGG	AGCTTCGGCG	GTACCAAGCG	CCCCGCGCTC	ATCGCGGTCA	180
CGGCGCGCGG	CGGCGGCAAC	GGCGGCAAGG	GGCGCGCGCG	TGGCAGCAAC	CCCAACGGGT	240
CAGGTGGCGA	CGGCGGCAAA	GGCGGCAACG	CGGTGCGCGG	CGGCAACGGG	GGCTCGATCG	300
GGGCGAACAG	CGGCATCGTC	GGCGGTTCGG	GTGGGCGCGG	TGGCGCTGGC	GGCGCGCGCG	360
GAAACGGCAG	C					371

## 2. INFORMATION FOR SEQ ID NO:238:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 424 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: cDNA

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

CTCGGGGTTC	CACCAAGCGG	CGGCGCGCGG	CTAGCGCGCG	GGCGCGACCA	CGCCCTTTTC	60
TTGACTCGTT	CAAGAAAAGG	CGCTTCTGTT	TGGTCGGCCA	TGTTGGCATT	ATCGTGACCC	120
ATGGGCAACA	TGACGTCGGA	ATCTCGCGCG	AAGGTCTAGC	TGCATCGGAA	TGGCGCGCGG	180
GGTGGTGAAT	ATGCTCTGAG	CGCTCATAGC	AGGGTTCTTC	GTACCTTTTC	CCGAGCGACA	240
CGCTCGGAG	CGGCGCGGTC	TCTGCTACGC	GGTCTCTGGA	AAGGGGTTCG	TGGGCAACAT	300
ATCGCGCGCG	CGCATCGGCT	GGGAGCGCGT	TTTCAAGAA	CGTTTCAAGG	CTTTTTCGCT	360
CGAATAGCG	CGCTCAAGAA	ATCGGCTGGA	ATCGCGGCTC	CTGAGGCTCT	AAGACGATCC	420
CGAC						424

## 2. INFORMATION FOR SEQ ID NO:239:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 317 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:239:

GCGATGGCGG	CCGCGGCTAC	CACCGCCAAT	GTGGAACGGT	TTCCCAACCC	CAACGATCCT	60
TTGCATCTGG	CGTCAATTGA	CTTCAGCCCC	GCCGATTTGG	TCACCGAGGG	CCACCGTCTA	120
AGGGCGGATG	CGATCCTACT	GCGCCGTACC	GACCGGCTGC	CTTTCGCCGA	GCCGCCGGAT	180
TGGGACTTGG	TGGAGTCGCA	GTTGCGCAG	ACCGTCACCG	CCGACACGGT	GCGCATCGAC	240
GTCATCGCCG	ACGATATGCG	TCCCGAACTG	GCGGCGGCGT	CCAAACTCAC	CGAATCGCTG	300
CGGCTCTACG	ATTCGTC					312

## (2) INFORMATION FOR SEQ ID NO:240:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 422 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (2) MOLECULE TYPE: cDNA

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:240:

TCCCGTATGC	GCTTCGCAGC	CGGTGCGCGC	TCAACGCGCC	GGAGGCAATC	GCTTCGCTGC	60
CGACGAATGG	TTGCATCAGC	ATCGCAGTGT	GCGGTCGTGC	ACCGACACCG	CGGTCCAACG	120
TGAAGTCAGC	GCGGAAATC	GGCGGAAATC	TGCGCCCTCAG	TTCAAGCTCG	CGCCCTAACG	180
GTTCTGGAAG	TTGGGTGCGC	GCTTCTCGGC	GAACGCGCGC	GGGCTTCCT	TGGCGTCGTC	240
GGACAGGAAG	ACCTTGATGC	CGATCTGGGT	GTGCATCTTG	AACGCTCGT	TTTCGGGCAT	300
CCACTCGGTC	TGCGGATGG	ACCGCAAGAT	GGCTTCGACG	GGCAGGGGTC	CGTTAGCCGA	360
GATGGCGCTC	GCAAGTTCTA	GAAGTTGGT	CAACGCTGCG	CGGTGGGGCA	CACGTGGCCC	420
AT						422

## (3) INFORMATION FOR SEQ ID NO:241:

## (1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 426 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

## (2) MOLECULE TYPE: cDNA

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:241:

ATTTCCTGCT	GAAGACGAGC	CGCGGGCTGT	TGATGCTGCT	CGATCTCGTA	GTCCCGGCGG	60
CGGCTGCTGT	CGTCTGCTCA	CGGGGCGCGG	GAAGGATAAG	GTCCCTAATG	CGCAGGTAGC	120
CGCCGAGGTC	CATGAGTGG	ATGATGATGC	GACTTCTGAG	GTCCCGGAGC	CGCAGCTTGG	180
GATCGGCGCT	GATCAGGCGC	GACGCGTAGG	ATAAGTGGAT	CGAATGCATA	GTGGCGTCCA	240
GATGCGCGCT	CGCACTTCCG	CGTCTCTCCA	CGCGAAATG	CTTGATTTCT	AGCTCCGCGT	300
ATTGTTTGGG	CATGCGCTGT	CGCATGAATG	CGAAATGCA	CATCGCGACA	AACGGGTCTG	360
AGCTGAGCTT	CGCGGCTTTC	CGCATGATGC	TGATATGCTG	GAATGCTGCT		420
ATGGA						426

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

AGACCGGCGA GGGTGTGGTC GCTGCCCGCG GCATTGTGCGA TAATCTGCGC TGGGTGCGACG	60
CGCCGATCAA CTAGTGAGGC GCAACGCTAG GCTTTGGGAT ACCCACAGCT AAAAAGTTTA	120
TCAAAGAAAC GAAGAAGGTT GCCATGAGCA CTGTTGCCGC CTACGCCGCC ATGTGCGCGA	180
CCGAACCCCT GACCAAGACC ACGATCACCC GTCGCGACCC GGGCCCGCAC GACATGGCGA	240
TCGACATCAA ATTCGCCCGA ATCTGTCGCT CGGACATCCA TACCGTCCAA ACCGAATGGG	300
GGCAACCGAA TTTACCTGTG GTCCCTG	327

(i) INFORMATION FOR SEQ ID NO:243:

1. SEQUENCE CHARACTERISTICS:

A. LENGTH: 125 amino acids  
B. TYPE: amino acid  
C. STRANDEDNESS: single  
D. TOPOLOGY: linear

ii. MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

Asp His Gly Gly Pro Ala Thr Asn Pro Gly Ser Gly Ser Arg Gly Gly	
1 5 10 15	
Ala Gly Gly Ser Gly Gly Asn Gly Gly Ala Gly Gly Asn Ala Thr Gly	
20 25 30	
Ser Gly Gly Lys Gly Gly Ala Gly Gly Asn Gly Gly Asp Gly Ser Phe	
35 40 45	
Gly Ala Thr Ser Gly Pro Ala Ser Ile Glu Val Thr Gly Ala Pro Gly	
50 55 60	
Ile Asn Gly Gly Lys Gly Gly Ala Gly Gly Ser Asn Pro Asn Gly Ser	
65 70 75 80	
Gly Gly Asp Gly Gly Lys Gly Gly Asn Gly Gly Ala Gly Gly Asn Gly	
85 90 95	
Ile Ser Ile Ile Ala Asn Ser Ile Ile Val Gly Ile Ser Gly Gly Ala	
100 105 110 115	
Ala Ala Gly Gly Ala Ile Gly Asn Ile Ser	
120 125	

(i) INFORMATION FOR SEQ ID NO:244:

1. SEQUENCE CHARACTERISTICS:

A. LENGTH: 104 amino acids  
B. TYPE: amino acid  
C. STRANDEDNESS: single  
D. TOPOLOGY: linear

```

Met Ala Ala Ala Gly Thr Thr Ala Asn Val Glu Arg Phe Pro Asn Pro
 1           5           10           15
Asn Asp Pro Leu His Leu Ala Ser Ile Asp Phe Ser Pro Ala Asp Phe
 20           25           30
Val Thr Glu Gly His Arg Leu Arg Ala Asp Ala Ile Leu Leu Arg Arg
 35           40           45
Thr Asp Arg Leu Pro Phe Ala Glu Pro Pro Asp Trp Asp Leu Val Glu
 50           55           60
Ser Gln Leu Arg Thr Thr Val Thr Ala Asp Thr Val Arg Ile Asp Val
 65           70           75           80
Ile Ala Asp Asp Met Arg Pro Glu Leu Ala Ala Ala Ser Lys Leu Thr
 85           90           95
Glu Ser Leu Arg Leu Tyr Asp Ser
 100

```

3) INFORMATION FOR SEQ ID NO:245:

1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2) MOLECULE TYPE: protein

3) SEQUENCE DESCRIPTION: SEQ ID NO:245:

```

Ala Tyr Ala Leu Arg Ser Arg Cys Arg Val Asn Ala Pro Glu Ala Ile
 1           5           10           15
Ala Ser Leu Pro Arg Asn Gly Ser Ile Thr Ile Ala Val Cys Arg Arg
 20           25           30
Ala Pro Thr Pro Pro Ser Asn Val Asn
 35           40

```

4) INFORMATION FOR SEQ ID NO:246:

1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 25 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

2) MOLECULE TYPE: protein

3) SEQUENCE DESCRIPTION: SEQ ID NO:246:

```

Val Pro Leu Asn Thr Ser Pro Arg Leu Pro Arg Leu Arg Asp Ser Val
 1           5           10           15
Ala Pro Val Ala Ser Leu Leu Ser
 20

```

- (A) LENGTH: 61 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

```

Met Ser Thr Val Ala Ala Tyr Ala Ala Met Ser Ala Thr Glu Pro Leu
 1             5             10             15
Thr Lys Thr Thr Ile Thr Arg Arg Asp Pro Gly Pro His Asp Met Ala
 20             25             30
Ile Asp Ile Lys Phe Ala Gly Ile Cys Arg Ser Asp Ile His Thr Val
 35             40             45
Gln Thr Glu Trp Gly Gln Pro Asn Leu Pro Val Val Pro
 50             55             60

```

(2) INFORMATION FOR SEQ ID NO:248:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

```

GCTTGGAGCC GTGGAGCCAC GGTGTGGGTG TGGGGGTGGA TTGCTTCTCG GCGAAAGTCA      60
ACTAAAGACC ACCTTGACAC GGAACGGGGC TCCCGCCATC GCGCGTCCGC GCTAGAGAGC      120
TTTGACCGCC GCGCGAAGCC TTGCTGTGTC TGGCGCATGC AGATCCGACA GCTTTCCTTC      180
AACATCGGGT GAGGCGGGTC GTAACGGGCG GCT
                                         213

```

(2) INFORMATION FOR SEQ ID NO:249:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

```

TTGAGCTGCT GTTGGGGGCT GCGGGTGGCA TGGGGGTGGG TGGGGCGGCG ATCGACGGCG      60
GACTTGCTGC TACCGCGCGG ACCTGGCGAT AAGGGCGAAT GCGGCGGAC GTTGGATGGC      120
TTTCAATGCG TGGGGTGGCT GCGGGTGGCT TTTAAATGTT TTTAATGTT GTTGGATGCG      180
ATGCTTGGCT GTTGGGTGCT ACGGGTGGCA TGGGGGTGAT TGGGGCGGCG GTTGGGGTGG      240

```

(A) LENGTH: 420 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:250:

AAGCGGTGAT	TGGCAAGGCG	ACCGCGCAGC	GGCGCGTAGC	CGCGGGACGG	CCCAGGCCCC	60
GACCGCAGCG	GCCTGTGTCT	GACCGGTCA	GGCACCAGCG	CGCGTGACCG	TCCCGCTCGT	120
CTACTTCGAC	GCCAGCGCCT	TGTCAAACT	TCTCACCAGC	GAGACAGGGA	GTCCTGTGGC	180
GTCCGCTCTA	TGGGACGGCT	GCGACGCCGC	ATTGTCCAAC	CGCGTGGCCT	ACCCCGAAGT	240
CGCGCCCGCA	CTCGCTGCAA	CGGGCCCGAA	TCACGACCTA	ACCGAATCGG	AGCTCGCCCA	300
CGCGGAGCGT	GACTGGGAGG	ACTTCTGGGC	GCGACCGCGC	GAGTCSAAGT	CACCGCCAGC	360
ATTGAACAGC	ACGCGCGGCA	GCTCGCGCGA	ACACATGCCT	TACGCGGAGC	CGACACCGTT	420

(A) LENGTH: 299 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:251

[illegible]

A LENGTH 1000000  
B TYPE 1000000  
C STRANDEDNESS single  
D TOPOLOGY linear

3. FREQUENCY OF OCCURRENCE OF THE TWO TYPES

## (2) INFORMATION FOR SEQ ID NO:253:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

Glu Leu Leu Phe Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly  
 1 5 10 15  
 Thr Asp Gly Gly Pro Gly Ala Thr Gly Gly Thr Gly Gly His Gly Gly  
 20 25 30  
 Val Gly Gly Asp Gly Glu Trp Leu Ala Pro Gly Gly Ala Gly Gly Ala  
 35 40 45  
 Gly Gly Glu Gly Gly Ala Gly Gly Ala Arg Ser Asp Gly Gly Ala Leu  
 50 55 60  
 Gly Gly Thr Gly Gly Thr Gly Gly Thr Gly Ala Gly Gly Ala Gly  
 65 70 75 80  
 Gly Arg Gly Thr Leu Leu Leu Gly Ala Gly Gly Glu Gly Gly Leu Gly  
 85 90 95  
 Gly Ala Gly Gly Glu Gly Gly Thr Gly Gly Gly Arg Arg Arg Trp Arg  
 100 105 110  
 Ser Gly Gly Cys Glu Trp His Trp Trp  
 115 120

## (2) INFORMATION FOR SEQ ID NO:254:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

Arg Val Leu Gly Leu Ala Thr Ala Glu Arg Pro Val Ala Ala His Arg  
 1 15  
 Arg Val Arg Val Glu Arg Pro Val Ser Asp Arg Val Ser Asp Glu  
 20 25 30  
 Arg Arg

## (2) INFORMATION FOR SEQ ID NO:255:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids

(11) MOLECULE TYPE: protein

(12) SEQUENCE DESCRIPTION: SEQ ID NO:255:

```

Leu Val Gly Gly Ile Gly Gly Thr Gly Gly Thr Gly Gly Asn Ala Gly
 1           5           10           15
Met Leu Ala Gly Ala Ala Gly Ala Gly Gly Ala Gly Gly Phe Ser Phe
 20           25           30
Ser Thr Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly Leu Phe
 35           40           45
Thr Thr Gly Gly Val Gly Gly Ala Gly Gly Gln Gly His Thr Gly Gly
 50           55           60
Ala Gly Gly Ala Gly Gly Ala Gly Gly Leu Phe Gly Ala Gly Gly Met
 65           70           75           80
Gly Gly Ala Gly Gly Phe Gly Asp His Gly Thr Leu Gly Thr Gly Gly
 85           90           95
Ala Gly Gly

```

(2) INFORMATION FOR SEQ ID NO:256:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(12) SEQUENCE DESCRIPTION: SEQ ID NO:256:

```

TCTGTTTGGG CGCCCGCGCG GTGGGCGGTT TTGGGCGGTGA CGGTGTGCCA TTCTGCGGCA 60
TCTGCGCGCG CGGGCGCGGT GTGCGCGGCG GGGCGCGGTGG CTGTTTCAGC CTCGGTGGGG 120
CGCGCGCGCG CGCCCGCAAT CGATTGCTGG GGAACAGCGG TCGCGCGCGG TCGCGCGCGGT 180
CGCGCGTGGT CTGGGCGGAC GCGCGTGGCG GCGCGCGCGG TGGGGTGGCG TCGACTACCG 240
CGCGTGGCGG CGCGCGCGCG GCGAACCGCA GCGTGTGGT AA 282

```

(2) INFORMATION FOR SEQ ID NO:257:

(1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 415 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(12) SEQUENCE DESCRIPTION: SEQ ID NO:257:

```

CGGACAGAGT CCGCTTATG GTCACTGAT GCGTJATG TGATTTAT GCGCTGAT 60
TATGCTGCTT CAGGATGGA GGAATATG AAAACCAATC TCGGCTGGA GTGAATAT 120
TATGCTGCTT CAGGATGGA GGAATATG AAAACCAATC TCGGCTGGA GTGAATAT 180

```

ATGCGTGGCAG CCCACTCGAC ACCACCGGTG GCGAACATCG AGGTCAACAC GCCGT

415

(2) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

TCACCGCGTG AACGGTTCGT AACACTGATA CGTATGCTTG TCAGCGAGCA GATCAAGTCC	50
AGTCCGACCA ATGCCAGGAG ATCATCGGCT AGGCTCACGG TTTCGCGTGG GACGAGACGG	100
CATTGAGTTC TGGCGTTGCA CGGTCCGTCC CTGGGTGGGA AGTCTGACGC CGCATCAGAA	150
TGGTTGTCAA TACCACTGTT TGGGGGATAT GGCCTATTTG GTGTGGTCCG GCGCGTCCAT	200
CGGATCCCTT TTCGAACGTT GCGCAAGCCC CGTCCAGTTA CGGCGTGTTC ACTGCGCGCT	250
CGCGTAGCTG CGCGGCGTCC ATCGGTTTGA ACCTCATCCC AATTCGCCCA ATGGGTGAGT	300
ACCTGACGCT GCT	373

(2) INFORMATION FOR SEQ ID NO:259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 403 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

CGAAACCCGA CAGGCCGGCA CGGACGGTCC GAAGTTGCAU AACGTTGGGC GCTCCATGTA	50
CGGAACCGGT CACCAACGGC TAGACAGGAG ATCTGTGGAT CGGCGCTTCC GTGTGGTCCC	100
CGCCGACTAC CGCGGGGGCG AACCGCAGGC ACCAAAGCAA CGGCATCGAT ACGGGGATCC	150
CGACTCGTGC CGAATTGAG GTGGTGGAG AAGGTTGCGG CGGACTCGA ACCCGGGTCA	200
ATGATTGAGT TTAACCGCT TACCAATAA TACATAAAG GTTGGGGGG TGTAAACGGC	250
TTTGAAGGCC TTTTCTCTG AAAGGAGTAA TATATCGCG ATAACTGGT ATAGTAGCGA	300
GAAGGCGGCG AAGGATGCG GTTCCGAAAT GTTGGTAAAT TTAATAGGG ATGGAAGCGT	350
TTT	403

(2) INFORMATION FOR SEQ ID NO:260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

AGTGGCCAGC	CGGTGGGCCA	ATGCATCCAG	GTCCCGGTAC	GTCAGCTGAC	CATCGGCCCA	60
ACTGACCCGC	ACCGAGTCAG	GCTGTGCCGC	AGCGATTTCG	GCGAACCGGG	TATGCACCCC	120
GGGTGCCGAC	GTGCTCACAT	CCGGCAGGCC	GGGTGCCGTC	GGATCGTGCT	CGCCGTCCAG	180
CAGAATGTCT	ACGTCCGCGA	GCGGCCGATC	CCACCGGCTG	ACCAAGCGCT	GTAACACAGC	240
CAGCACCCGC	CTGCCGAGGC	TTTCGGGGCG	CATCGTGCCC	AGCGCACCGT	CGAGCACCTC	300
CACTAGCAGC	GTGACCTCAC	CGGTGCTGCG	GTGCGCGGCG	ACGGTCACCG	GAAAGTGCGA	360
CAAACCTCTT	AGCGCCACCG	GACGGAACCT	CACCCCGTTT	GCGA		404

## (2) INFORMATION FOR SEQ ID NO:261:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 421 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

11. MOLECULE TYPE: cDNA

## 12. SEQUENCE DESCRIPTION: SEQ ID NO:261:

GTCTGTCTCG	CAGGTGTGTC	ITCGAACCCG	CTGGTAACT	TGCGACCCCG	GTATCGGCCC	60
AGCATGGAAC	CGCGCAACG	GGCGGTCTCA	CGCGGTACTT	CGCAAGACCC	GGCGGTGCA	120
GTCCGACCAC	TGAGCGGCGA	CGCCCGGCGC	GCACTATTTC	ACAACGGCAC	CGCCCAATTG	180
GTCCCTCTGC	GCCCGGTCGC	CGATTCCGCG	GCACCGCGCA	GCATCATGGT	CTTCGATGAC	240
ATCCACCTTG	CACCGCTCGT	CATTTTCTG	CGCGCGCGCG	CAGCCCGGTT	GACCAGCGAC	300
GACGACCGCA	CGGCTTCTCT	TGCGCGCCCG	GGCGGCTACT	TGCTGGCGCA	CGTGTCTCTC	360
GTTCACACCG	CACGAGTGA	TGTCGCTGAC	GCAGCGCACA	CGGATTTCAC	CGCGATCGCC	420
						421

## 2. INFORMATION FOR SEQ ID NO:262:

## (1) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 426 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

11. MOLECULE TYPE: cDNA

## 12. SEQUENCE DESCRIPTION: SEQ ID NO:262:

ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	60
ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	120
ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	180
ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	240
ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	300
ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	360
ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	420
ATGATATTA	ATGTAAGCC	TATCGCTGCT	GCATCTTCT	TCGCGGCGCG	TAATTAATA	426

## 3. INFORMATION FOR SEQ ID NO:263:

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:263:

GTCTGGTTCG CAGGCTGTTT TTCGAACCCG CTGGCTAACT TCGCACCCGG	GTATCCGCCC	60
ACCATCGAAC CCGCCCAACC GGCGGTGTCA CCGCCTACTT CGCAAGACCC	GGCCGGTGCA	120
GTGCGACCAC TGAGCGGCCA CCCCCGGGCG GCACTATTTCG ACAACGGCAC	CCGCCAATTG	180
GTGGCTCTGC GCCCGGGGCG CGATTGGGCG GCACCCGCCA GCATCATGGT	CTTCGATGAC	240
GTGCACGTTG CACCGCCCGT CATTTTTCTG CCGGGCCCCG CAGCCGCGTT	GACCAGCGAC	300
GACCACGGCA CGGCCCTTCT TGCCGCCCCG GCGGGCTACT TCGTGGCCGA	CCTGTCTCTC	360
GGTCACACCG CACGAGTGAA TGTGCTGAC GCAGCGCACA CCGATTTCAC	CGCGATCGCC	420
CGCCGCTCCG ACGGCAAGCT GGTGCTGGGT AGCGCAGATG GCGCCGTCTA	CACGCTTGGC	480
AAGAACCCTG AGTTGACCCG CGTCGGCGCC GCCACCGTAG CC		522

2. INFORMATION FOR SEQ ID NO:264:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 739 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:264:

CTTGGGGGCG ACCGGCTGCT GCGGGCCCCA CCGCCTGGGG CGAGACCCCG	CGCAAAACCA	60
ACCGGTGGCC GTTAGTGGCT GCGGGCCCCG TGCTGCTGCT GTGCTCTCTG	TTGGGGCGCCA	120
TGGGCATCTG CATGCGCATT TGGGGCAAGC CGGTACAGGC GGTTCAGCCC	GTTCGGGAGG	180
AGCGGCTTAG CCGCCTACTG CTGAAGTCTT CAGAAGTCAA TGGCGTGATG	GGTCTGCTCT	240
CGTGCAGCC TGGCAAAAGC ATGACATCGA TGGACTCTTG GCGGCTGACC	GTGTCCCTGG	300
TGGACTGCCA GGGGCGCCCT TATACAGCC AGGATCGGCT GTATCGCGCC	ACCGGCTACA	360
CGGCCATCAA CCGCTTGATT TATCGGACC CCGGCGACAA CTACGAACAT	TGGGTGAACC	420
AAGCGCTCTT CCGCTTTCCG ACCGGCGACA AAGCCCGCGG GTTCGTGCAG	ACTTCGGGCC	480
ACAAATGGAA GAAGTCCGCA TCGAAGACCG TCACCGTCAC GAATAAGGCG	AAGACCTACC	540
GTGTGACGTT TCGCGACGTT GAAGGCAAGT TCGGACGAT TACGGTGATA	GACACCGAAT	600
CAGGCTCTGA CGGTGGGAA TGTAAAGGCT CATGAGGCT TCGCAAGAT	GTGTTCTCTG	660
AGTGAAGCT ATCGGCTA TATTAAGTA ATGAGCAGG TATATCTCT	TATAGATCTT	720
TTTAAAGG TTAAGG		739

2. INFORMATION FOR SEQ ID NO:265:

1. SEQUENCE CHARACTERISTICS:

A LENGTH: 63 base pairs  
B TYPE: nucleic acid  
C STRANDEDNESS: single  
D TOPOLOGY: linear

AGACGTCGTC GAGGCCGCCA TCGCCCGCCG CCAAGCCGTT AACCCGGCAC TGAACCCGTT 60  
GGCGTATGC 69

(2) INFORMATION FOR SEQ ID NO:266:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 523 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:266:

ACTGCACCCG GCAGGCCGCA CCAACCGATC GGTGCACTA GCACTGCCCG TGGAGGCCCC 60  
CCCGCGGTCT GTGCTTCCG ACCGGGAACG CTGCGGCAGC GCGGCTCCAG AAGGGTTGGA 120  
GGGAGAGTTG GACGACCTA TCGACGAGCG GTTCCCGGTC TTCAGCTCCG CCAGTCTCCG 180  
CGAAGCGCTG CCGGGTCCCG TGACCCCGAT GACGCTGGAT CTCGAGTTGA CTCGACTGCC 240  
CGCGGCCCGT CCGGCGATCG CTCGGCTACT GCGGCTTGGC GGTGTCGTTG CCGATGAGTG 300  
GGAGAGAAGA GGCATCGCGG TGTTCGCTCA CCGCCCGTAT ATCGGAGTGT CCGCCAATAT 360  
TGTGCGCGCC GCCCAACTGC CCGGGTGGGA CCGCGAGGCC GTAACCCCGC GGGCACTGGG 420  
CGAGCAACCG CAGGTCACG AGCTGCTTCC GTTGGTCTGA CCGCAACTTG CCGGCCGACC 480  
GCTCGGCTCG GTCGCGAAGG TGGTCTGAC GGCACGGTCG CTG 523

(2) INFORMATION FOR SEQ ID NO:267:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 214 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:267:

TTTCTGGTGT CTTGGGGGTA GGAGCGACTT CCGCCCGCCG CCGCGGCGGT GGAGCGGGGT 60  
TTGAGGAGAG CCGTCCCGGC TCGGCGGCTT TCGGACCAAA AGGCGCGATC TATTGCGGCA 120  
TTTCTGGTGT CCGCGGCTT TCGGCGGCTT TCGGCGGCTT TCGGCGGCTT TCGGCGGCTT 180  
TTTCTGGTGT CCGCGGCTT TCGGCGGCTT TCGGCGGCTT TCGGCGGCTT TCGGCGGCTT 240

(2) INFORMATION FOR SEQ ID NO:268:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 521 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

TGAACTGCTC	TTGGCCATAG	CGGGCCCTTG	TCTCGGCGCTT	GTCCAAACCC	TGCAGCGCGC	120
CGTAGTGGCG	TTGGTTGAGC	CGCAGGTAC	CGCGCACGGG	AATCCAGAGC	CGATCGGCGC	180
TGTCCAACGC	CAGATGCGCG	GTGGTGATCG	CGCGCGCGAG	CAACGAGGTG	TAGAGCACGT	240
CGGGCAATAG	GTCTGTGTTCC	CGGATCAGCT	CGCGGCTTCC	AACCGCCTCT	GCCTGGCCCT	300
TGTCCGTGAG	GGCGACATCG	ACCCAGCCCG	TGAACAGGTT	GAGGGCATTG	CAGTCGCTCT	360
CGCCGTGGCG	CAGCAACACC	AGGCTGCCAG	TGTTTGCCAT	ACCGGCAAGT	CTCTCACGCA	420
CTCCCGCACT	CCTCATCGTG	GACCAAAATG	CCCGAATTCT	CCTCGGTCCG	CTGCGCAGCG	480
CGTTCATACC	GGCGAGGIGG	TCGGCACCGT	AACGGCCCGT	T		521

## (2) INFORMATION FOR SEQ ID NO:269:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (2) MOLECULE TYPE: cDNA

## (3) SEQUENCE DESCRIPTION: SEQ ID NO:269:

CTCCAGGCTC	ATTGGTCTGA	ACAAAGCCAC	CGGGCCCTAG	AGCGGACGCG	CCCATTCGTT	60
CTCGTGTATG	TGGCGGTACA	GCTGGGCATC	GGGCGCTGGA	CGAACCTGCG	CCGAGGCGCA	120
CGGAACGAGC	CGGTGCTGGC	TCACCCGGGG	TCAGAACGGT	AGTGCCAGAC	AGTCTCGCGC	180
CGCGAAGGGT	TTGACGGGTC	AGACTCGGCC	TGCGGCTGTT	CGACCGAGGC	GTGGATCGCG	240
CGGAGCTGAG	AGCGTAGCGC	CTCGAAGTCA	CGGCGGAGCC	GTTCGAGCAC	CGAGTCCACC	300
TGCGTGGTCT	TGTTCCCGCG	CAGCACCTGC	GTGAAGTTGA	CGCGGTGCGC	ATCGGCGCGC	360
GTGACCGCGA	ACGCGCGGAG	CGTCTCGGCC	GTGCTCGGCC	GGCGCAGGCG	CGGCAAGTGC	420
TGCGCA						426

## (2) INFORMATION FOR SEQ ID NO:270:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (2) MOLECULE TYPE: cDNA

## (3) SEQUENCE DESCRIPTION: SEQ ID NO:270:

CTGAAGGCG	CGGAAGGCG	CGAATGCG	CTGAGGCG	CTGAGGCG	CTGAGGCG	60
CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	120
CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	180
CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	240
CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	300
CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	360
CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	CGGAGGCG	420

## (2) INFORMATION FOR SEQ ID NO:271:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 111 base pairs

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

```

AAGATCATCG GCGCCCTCC TTAGCATCGC TCGCTCTGTC ATCGTCGCGG GCGCGGATCA      60
CGGAGGTCCG GCCTTGTAAC CCACTCCTCG AACGGTCAGC ACCACAGTCG GGTTCGCGG      120
ATCCTTTTCG ACCTTGGCCC GCAGACGCTG GACATGCACG TTCACCAGCC TGGTATCGGC      180
TGGGTGCCGG TAACCCCATC CCGTTCGAG CAGCACATCA CGAGTAAACA CCTGGCGCGG      240
CTTGCGCGCC AATGCGACCA ACAGGTCGAA TTCCAGCGGT GTCAACGAGA TCTGCTCACC      300
GTTGCGAGTG ACCTTGTCG CCGGTACGTC GATTTCTACG TCGGCGATGG ACAGCATCTC      360
GCGGGGTTTC TCGTCGTTGC GCGCGAGCCG CGCCCGCACC CGCGCAACCA GCTCCTTGGG      420
CTTGAACGGC TTGATGATGT AGTCGTCGCG GCGCGACTCC AGACCCAGCA CCACATCCAC      480
GGTGTGCGTC TTGCGGTGA GCATCAGGAT CGGAACACCG GAATCGGCGC GCAACACCCG      540
GCACACGTCG ATGCGGTTC TACCGGGGCA A

```

(i) INFORMATION FOR SEQ ID NO:272:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 93 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

```

Leu Phe Gly Ala Gly Gly Val Gly Gly Val Gly Gly Asp Gly Val Ala      1
1      5      10      15
Phe Leu Gly Thr Ala Pro Gly Gly Pro Gly Gly Ala Gly Gly Ala Gly      20
20      25      30
Gly Leu Phe Ser Val Gly Gly Ala Gly Gly Ala Gly Gly Ile Gly Leu      35
35      40      45
Ala Gly Asn Ser Gly Ala Gly Gly Ser Gly Gly Ser Ala Leu Leu Trp      50
50      55      60
Gly Asp Gly Gly Ala Gly Gly Ala Gly Gly Val Gly Ser Thr Thr Gly      65
65      70      75      80
Gly Ala Gly Gly Ala Gly Gly Asn Ala Ser Leu Leu Val
85

```

(i) INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

20

25

## (2) INFORMATION FOR SEQ ID NO:274:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

Lys Pro Asp Arg Pro Ala Ala Thr Val Gly Ser Cys Thr Thr Val Arg  
 1 5 10 15  
 Ala Pro Cys Ser Gln Pro Val Thr Thr Ala  
 20 25

## (2) INFORMATION FOR SEQ ID NO:275:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

Pro Pro Ala Gly Arg Pro Met His Pro Ala Pro Gly Thr Ser Ala Asp  
 1 5 10 15  
 His Pro Pro Asn  
 20

## (2) INFORMATION FOR SEQ ID NO:276:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 190 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

His Leu Val Ala Gly Cys Ser Ser Asn Pro Leu Ala Asn Thr Ala Pro  
 1 5 10 15  
 Lys Pro Gly Thr Ser Pro Ala Gln Thr Val Val Ser Ser Thr  
 20 25

```

      50              55              60
Pro Gly Ala Asp Ser Ala Ala Pro Ala Ser Ile Met Val Phe Asp Asp
65              70              75              80
Met His Val Ala Pro Arg Val Ile Phe Leu Pro Gly Pro Ala Ala Ala
      85              90              95
Leu Thr Ser Asp Asp His Gly Thr Ala Phe Leu Ala Ala Arg Gly Gly
      100             105             110
Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val
      115             120             125
Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala
      130             135             140

```

(x) INFORMATION FOR SEQ ID NO:277:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

```

Met His Ile Thr Leu Asn Ala Ile Leu Arg Ala Ile Phe Gly Ala Gly
 1              5              10              15
Gly Ser Glu Leu Asp Glu Leu Arg Arg Leu Ile Pro Pro Trp Val Thr
      20              25              30
Leu Gly Ser Arg Leu Ala Ala Leu Pro Lys Pro Lys Arg Asp Tyr Gly
      35              40              45
Arg Leu Ser Pro Trp Gly Arg Leu Ala Glu Trp Arg Arg Ala Tyr Asp
      50              55              60
Thr Val Ile Asp Glu Leu Ile Glu Ala Glu Arg Ala Asp Pro Asn Phe
      65              70              75
Ala Asp Arg Thr Asp Val Leu Ala Leu Met Leu Arg Ser Thr Tyr Asp
      80              85              90              95
Asp Gly Ser Ile Met Ser Arg Lys Asp Ile Gly Asp Glu Leu Leu Thr
      100             105             110
Leu Leu Ala Ala Gly His Glu Thr Thr Ala Ala Thr Trp Ala Gly Arg
      115             120             125
Asn Gly Ser Thr Gly Thr Arg Thr Val Ser Arg Leu Trp
      130             135             140

```

(x) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 161 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```

Val Leu Val Ala Gly Cys Ser Ser Asn Pro Leu Ala Asn Phe Ala Pro
 1           5           10           15
Gly Tyr Pro Pro Thr Ile Glu Pro Ala Gln Pro Ala Val Ser Pro Pro
          20           25           30
Thr Ser Gln Asp Pro Ala Gly Ala Val Arg Pro Leu Ser Gly His Pro
          35           40           45
Arg Ala Ala Leu Phe Asp Asn Gly Thr Arg Gln Leu Val Ala Leu Arg
          50           55           60
Pro Gly Ala Asp Ser Ala Ala Pro Ala Ser Ile Met Val Phe Asp Asp
          65           70           75           80
Val His Val Ala Pro Arg Val Ile Phe Leu Pro Gly Pro Ala Ala Ala
          85           90           95
Leu Thr Ser Asp Asp His Gly Thr Ala Phe Leu Ala Ala Arg Gly Gly
          100          105          110
Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val
          115          120          125
Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala Arg Arg Ser Asp
          130          135          140
Gly Lys Leu Val Leu Gly Ser Ala Asp Gly Ala Val Tyr Thr Leu Ala
          145          150          155          160
Lys Asn Pro

```

## (2) INFORMATION FOR SEQ ID NO:279:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (iii) SEQUENCE DESCRIPTION: SEQ ID NO:279:

```

Trp Gly Ala Pro Pro Ser Gly Gly Pro Ser Pro Trp Ala Gln Thr Pro
 1           5           10           15
Arg Lys Thr Asn Pro Trp Pro Leu Val Ala Gly Ala Ala Val Val
          20           25           30
Leu Val Leu Val Leu Gly Ala Ile Gly Ile Trp Ile Ala Ile Arg Pro
          35           40           45
Leu Ile Val Gln Ser Thr Val Pro Val Val Ala Gly Arg Leu Ser Val
          50           55           60
Ser Leu Leu Asn Ser Ser Ser Val Asn Ala Val Met Gly Ser Ser Ser
          65           70           75           80
Met Gln Pro Gly Leu Pro Ile Thr Ser Met Asp Ser Ser Pro Val Thr
          85           90           95
Val Ser Leu Pro Asn Lys Gln Gly Ala Leu Tyr Thr Ser Gln Asp Pro
          100          105          110
Val Thr Ala Gly Thr Thr Thr Thr Ala Ile Asn Gly Leu Ile Ser Ser
          115          120          125
Val Thr Thr Ser Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
          130          135          140
Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
          145          150          155          160
Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
          165          170          175          180
Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
          185          190          195          200
Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
          205          210          215          220
Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
          225          230          235          240
Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr

```

```

Lys Trp Lys Asn Cys Ala Gly Lys Thr Val Thr Val Thr Asn Lys Ala
      165              170              175
Lys Thr Tyr Arg Trp Thr Phe Ala Asp Val Lys Gly Ser Pro Pro Thr
      180              185              190
Ile Thr Val Ile Asp Thr Gln Glu Gly Ala Glu Gly Trp Glu Cys Gln
      195              200              205
Arg Ala Met Ser Val Ala Asn Asn Val Val Val Asp Val Asn Ala Cys
      210              215              220
Gly Tyr Gln Ile Thr Asn Gln Ala Gly Gln Ile Ala Ala Lys Ile Cys
      225              230              235              240

```

## (2) INFORMATION FOR SEQ ID NO:280:

## 1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## 11. MOLECULE TYPE: protein

## xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

```

Asp Val Val Glu Ala Ala Ile Ala Arg Ala Glu Ala Val Asn Pro Ala
 1              5              10              15
Leu Asn Ala Leu Ala Tyr
      20

```

## 2. INFORMATION FOR SEQ ID NO:281:

## 1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## 11. MOLECULE TYPE: protein

## xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

```

Asp Ala Pro Val Ile Lys Lys Lys Lys Ser Gly Lys Lys Ala Leu Arg
      1              5              10              15
Lys Ile Lys Arg Pro Arg Ser Val Ser Ser Ala Gly Lys Pro Leu Ala
      20              25              30              35
Ser Ala Ala Pro Lys Gly Lys Lys Lys Lys Lys Lys Asn Asp Arg Ile Asp
      40              45              50              55
Lys Arg Phe Pro Val Phe Ser Ser Ala Ser Leu Ala Glu Ala Leu Pro
      60              65              70              75
Gly Pro Lys Thr Pro Met Thr Lys Asp Val Lys Lys Ser Gly Leu Arg
      80              85              90              95
Lys Ala Lys Lys Ala Met Lys Lys Lys Lys Lys Lys Lys Lys Val Val
      100             105             110             115

```

```

      115              120              125
Trp Asp Ala Gln Ala Val Thr Arg Arg Ala Leu Gly Glu Gln Pro Gln
      130              135              140
Val Thr Glu Leu Leu Pro Phe Gly Arg Pro Gln Leu Ala Gly Gly Pro
      145              150              155              160
Leu Gly Ser Val Ala Lys Val Val Val Thr Ala Arg Ser Leu
      165              170

```

## (2) INFORMATION FOR SEQ ID NO:282:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

ii MOLECULE TYPE: protein

xi SEQUENCE DESCRIPTION: SEQ ID NO:282:

```

Val Gly Val Val Gly Val Gly Ala Thr Ser Pro Ala Gly Ala Gly Ala
 1             5             10             15
Gly Ala Gly Ser Ala Gly Thr Gly Ala Gly Ala Gly Gly Gly Ala Thr
      20             25             30
Lys Gly Arg Ile Asp Ser Ala Ser Ala Leu Ala Ala Pro Leu Ser Thr
      35             40             45
Gly Leu Leu Ala Val Pro Ser His Thr Thr Asn Gln Arg
      50             55             60

```

## (2) INFORMATION FOR SEQ ID NO:283:

## (i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 133 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

ii MOLECULE TYPE: protein

xi SEQUENCE DESCRIPTION: SEQ ID NO:283:

```

Tyr Ala Asn Thr Glu Leu Val Leu Leu Arg His Thr Glu Ser Asn
 1             5             10             15
Glu Asn Ala Ser Asn Ser Phe Thr His Trp Val Asn Val Glu Leu Thr
      20             25             30
Asn Lys Gly Gln Ala Glu Ala Val Arg Ser His Glu Leu Ile Ala Glu
      35             40             45
His Asp Leu Leu Pro Asp Val Leu Tyr Thr Ser Leu Leu Arg Arg Ala
      50             55             60
Glu Thr Thr Ala Val Leu Ala Leu Asp Ser Ala Asn Arg Leu Trp His
      65             70             75

```

Phe Met Ala Trp Arg Arg Ser Tyr Asp Thr Pro Pro Pro Pro Ile Glu  
 115 120 125  
 Arg Gly Ser Gln Phe  
 130

## (2) INFORMATION FOR SEQ ID NO:284:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 63 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

Pro Gly Ser Phe Ala Arg Thr Cys Pro Pro Gly Arg Thr Ala Asp Ala  
 1 5 10 15  
 Pro Ile Arg Cys Arg Asp Ser Arg Gly Thr Ala Gly His Arg Ala Leu  
 20 25 30  
 Asp Glu Pro Pro Pro Arg Gly Ser Glu Pro Ala Arg Arg Arg Ser Arg  
 35 40 45  
 Gly Val Arg Thr Val Val His Asp Ser Leu Ala Ala Arg Arg Val  
 50 55 60

## (2) INFORMATION FOR SEQ ID NO:285:

## (i) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 70 amino acids
- B) TYPE: amino acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

Gly His Gly Gly Gln Ser Ala Ile Gly Leu His Gly His Ala Gly Gly  
 1 5 10 15  
 Asn His Gly Gln Gly Gly Ala His Arg His Leu Trp His His Gly His  
 20 25 30  
 Val His Gly His Gly Gly Ala Asp Arg Thr Tyr Arg His Arg His Ala  
 35 40 45  
 Ala Arg Ser Gly Arg His Gly Arg Arg Gly Trp Arg Arg Trp Ala Asp  
 50 55 60  
 Arg Gln Arg Arg Gly Arg Arg Arg  
 65 70

## (2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

```

Asp His Arg Arg Arg Ser Leu Ala Ser Leu Arg Ser Ala Ser Ser Pro
 1           5           10           15
Ala Arg Ile Thr Glu Val Arg Pro Cys Thr Pro Leu Leu Glu Arg Ser
 20           25           30
Ala Pro Gln Ser Gly Ser Arg Asp Pro Phe Arg Pro Trp Pro Ala Asp
 35           40           45
Ala Gly His Ala Arg Ser Pro Ala Trp Tyr Arg Leu Gly Ala Gly Asn
 50           55           60
Pro Ile Pro Val Arg Ala Ala His His Glu
 65           70

```

2. INFORMATION FOR SEQ ID NO:287:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

```

CCGCGACGTAA CACCCTGAAT TGAAGGGAGG CCGTGGTCAT GGGCCGATTC TATCCGTGGG      60
CGAACGGTTA TTGACGGCCC GGAGGCGACT CCGCTGCGAC CAGTGGTCA CTCAGCGCGT      120
TTTCACGGCA ACGAAGCGGC GACATACCA TTGACATTC ACAGCACGGC CCCC      174

```

2. INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 104 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

```

TCCCAACGCG GTTGACCTTC CCGCGGCTCG CCGTAAAGAG TTCTCGCAC TTTCGGGTGA      60
CCCTCGCCGC GCACCGGAGG ACCGCTGAGC TCAAGCTGCT AGTGGAGGTG CTCGACGGTG      120
CCCTGGGCAC GATGGGCTTC GAAAGCGTCC GCAGCGGGGT GCTGGCTGTG TTAAGCGCT      180
CGCTCAGCGG CCGGATTCCT CCGCTGCGCC ACCTCGACAT TCTGCTGGAG TGGAGGAGAG      240
ATTCAGCCCG ATCGGCTCTT CCGGATTCCT CCACTTGGCT ACCGCGGGTG TATAGCGAGT      300
TCCCGGAAAT TCGTGGCTTC CAGCTGACT CCGTGGCTCT CAGTGGCTCT CAGTGGCTCT      360

```

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

```

Ala Asn Gly Val Thr Phe Arg Pro Val Ala Leu Glu Ser Leu Ser His
 1             5             10             15
Phe Pro Val Thr Val Ala Ala His Arg Ser Thr Gly Glu Leu Thr Leu
 20             25             30
Leu Val Glu Val Leu Asp Gly Ala Leu Gly Thr Met Ala Pro Glu Ser
 35             40             45
Leu Gly Arg Arg Val Leu Ala Val Leu Gln Arg Leu Val Ser Arg Trp
 50             55             60
Asp Arg Pro Leu Arg Asp Val Asp Ile Leu Leu Asp Gly Glu His Asp
 65             70             75             80
Pro Thr Ala Pro Gly Leu Pro Asp Val Thr Thr Ser Ala Pro Ala Val
 85             90             95
His Thr Arg Phe Ala Glu Ile Ala Ala Ala Gln Pro Asp Ser Val Ala
100             105             110
Val Ser Trp Ala Asp Gly Gln Leu Thr Tyr Arg Glu Leu Asp Ala Leu
115             120             125
Ala Asp Arg Leu Ala Thr
130

```

## 2. INFORMATION FOR SEQ ID NO:290:

## 1. SEQUENCE CHARACTERISTICS:

- A LENGTH: 325 base pairs  
 B TYPE: nucleic acid  
 C STRANDEDNESS: single  
 D TOPOLOGY: linear

## 11. MOLECULE TYPE: cDNA

## 11. SEQUENCE DESCRIPTION: SEQ ID NO:290:

```

ATTGTAATG ATAGCAATAC GTGTTCTGGG TGGTTTGTGG AGGTGCTAT GAGGACAAAG 60
TAAAGAAGA TACCAAGGAG GTGCGCGAGG TGTTCGCGGT TGGCGCGGTT AAATATTTG 120
TGTGGGGGG TGGGAAAGAG TCGACCGGCG ACTCGCGCGG AGCGTCGCGG AACGAGTTGG 180
TGTTCAGAG GTGGGAGAA CAGCGCGCTG AGACCGTGGG CGTTTGTGTT GAGGTTGTGG 240
AGAGGTGTA TCGCAAGATG GTGTCAGCT GCGCGCACTG GTTCAACAGG ATCGGCAAGG 300
AATATCGGA TGTGGGCGCG AACTACACCG TGCTGCACCA CAGCGAGCTG GTCAATCGGT 360
TGTGCGCGA TAAAGAGGTT GTGCGTGTCA GTCGGTTTG TCAGGACATG AGGTACCAAG 420
ACCGTGTGA TGTGCGTGGG GAGGACAAAG TCTACGAGGC ACCACGGGAG GTGATCGGTT 480
TGGCGCGGGT TACTGATGCG GAGATCGGCG GGCATGCCGA TCGGCG 540

```

## INFORMATION FOR SEQ ID NO:291:

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

CTCGCCGCCG	TGATCTGGCC	GGCGAACTTC	GTCAGTGCAT	CCAGACCCCA	ACGATCATCG	60
ATCAGGCCGA	TGCCCATGAT	CACCGCACCG	GCCACCAGCA	CCGCGGGCAT	GCCGGTGGAA	120
TAGACGAACC	CCCGGGTGAG	TGCCGGAAGC	TGGGAGGCAA	GAAAGACGGC	GCCGACAATG	180
CCCAGGAACA	TCGCCAACCC	ACCCATCCGA	GGGGTAGGCG	TGACGTGCAC	ATCTCGCTCC	240
CCCGGGTAGG	CGACCGCTCC	CAGGCGACTG	GCCAGCATCC	GCACCGGACC	GGTCGCAAAA	300
TAGGTGATGA	TCGCCCGCGT	CAGCCCCGAC	AGCGCAAGCT	CACGCAGCGG	GACACCGCGG	360
CCCGCATAGG	ACAGGGCGAG	CAAGCCACCG	GCAACGCCGG	CCACATCGCT	GGACACCTCG	420
AGACCGTACT	GCACCAACCT	GAAGAGCTGA	ACAATCGCGG	AACGTGCAAC	AGCTGCCAAC	480
AATTGGG						497

2. INFORMATION FOR SEQ ID NO:292:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 528 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

ACCAAGCCCG	ACAATATGAG	CCGGGGCAAC	CCCGCATGTA	CGAGCTTGAG	CTCCCGGCCG	60
CTCAGCTCTC	CTCCTCCGAC	AGCCGTGGTC	CCGTGTTGGT	GCACGCTTTC	GAAGGTTTCT	120
TCGAACCCCG	CCATCCGATC	CGGCTGGCGG	CCCGCCACCT	CAAGGCGGCG	CTGGACACAG	180
AGCTGCTGCC	CTCCTTCCCG	ATCGATGAAC	TACTGGACTA	CGGCTCGCGG	CGGCCATTAA	240
TGACTTTGAA	GACCGATCAT	TTGACCAACT	CCCATGATCC	TGAGCTAAGC	CTGTATCCGC	300
TGCGCGACAG	CATCGGCACC	CCATTCTCTG	TGCTGCGCGG	TTTGGAGGCG	GACCTGAAGT	360
TGGAGCCGTT	CATCACCGCC	CTCGGATTGC	TGCCCCAGCG	CTTGGGTGTA	CGGCAGAACG	420
ATCGGCTTGG	GCACCGTCCC	GATGGCCGTT	CCGCACACAC	GACCGATCAC	GATGACCGCT	480
CATTGCAACA	ACCGGGAGCT	ATCTCGGATT	TTCAACCGTT	CCATCTCG		528

3. INFORMATION FOR SEQ ID NO:293:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 510 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

```

CCCCCTTGCC GCGGCTGGTG GGGGTGGCGG CCTGGTTGAC GTATTGTTCC ACCGGCCCCGG 240
CCCTTGACCC TTGGGCGGTG TCGATCGCGG CGTCGATGGA TCCGCCGACC ACGACGTGCG 300
AAGCCTCGCC TGCCGCGGCA GCGGCCAAC TGTGTGCGCG CTCCTGCGAT TTGGCCCCGG 360
CCGACGAGAT GATGGGCACC ACCGAGCCT GCGGCCGTCT GGGGGAGGCC AGCGCGGGTT 420
CGCGGTCACG CCATACGCGA CGGTGCGCGG CCGCTTCGGA GATTTGCAGG CTGCGTTGCA 480
CCAGATCGAG CAGCGGTCTG CCCAGGGACT GGGTTAGCCC GTTGGCGCCG CCGTTGTAGC 540
GGCGAGCGCA ATATCGGTGC CCACTCGACC CAACCGCGAC TCCATAAGCG ACACCATTGG 600
CGGTTGATGC
610

```

## (2) INFORMATION FOR SEQ ID NO:294:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 164 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: protein.

## (3) SEQUENCE DESCRIPTION: SEQ ID NO:294:

```

Phe Asp Gly Tyr Glu Tyr Leu Phe Trp Val Gly Cys Ala Gly Ala Tyr
1      5      10      15
Asp Asp Lys Ala Lys Lys Thr Thr Lys Ala Val Ala Glu Leu Phe Ala
20     25     30
Val Ala Gly Val Lys Tyr Leu Val Leu Gly Ala Gly Glu Thr Cys Asn
35     40     45
Gly Asp Ser Ala Arg Arg Ser Gly Asn Glu Phe Leu Phe Gln Gln Leu
50     55     60
Ala Gln Gln Ala Val Glu Thr Leu Asp Gly Leu Phe Gln Gly Val Glu
65     70     75     80
Thr Val Asp Arg Lys Ile Val Val Thr Cys Pro His Cys Phe Asn Thr
85     90     95
Leu Thr Lys Glu Tyr Arg Gln Leu Ile Ala Asn Tyr Thr Val Leu His
100    105    110
His Thr Gln Leu Leu Asn Arg Leu Val Arg Asp Lys Arg Leu Val Pro
115    120    125
Val Thr Pro Val Ser Gln Asp Ile Thr Tyr His Asp Pro Tyr Tyr Leu
130    135    140
His Arg His Asn Lys Val Thr Glu Val Arg Glu Leu Thr Glu Ala
145    150    155    160
Gly Gly Ala Thr

```

## (2) INFORMATION FOR SEQ ID NO:295:

## (1) SEQUENCE CHARACTERISTICS:

- A LENGTH: 161 amino acids
- B TYPE: amino acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

Arg Arg Arg Asp Leu Ala Gly Glu Leu Arg Gln Cys Ile Gln Thr Pro  
 1 5 10 15  
 Thr Ile Ile Asp Gln Ala Asp Ala His Asp His Arg Thr Gly His Gln  
 20 25 30  
 His Arg Gly His Ala Gly Gly Ile Asp Glu Pro Pro Gly Glu Cys Arg  
 35 40 45  
 Lys Leu Gly Gly Lys Lys Asp Gly Ala Asp Asn Ala Gln Glu His Arg  
 50 55 60  
 Gln Pro Thr His Pro Arg Gly Arg Arg Asp Val His Ile Ser Leu Pro  
 65 70 75 80  
 Arg Val Gly Asp Gly Ser Gln Ala Thr Gly Gln His Pro His Arg Thr  
 85 90 95  
 Gly Arg Lys Ile Gly Asp Asp Arg Arg Gly Gln Pro Asp Gln Arg Lys  
 100 105 110  
 Leu Thr Gln Arg Asp Thr Gly Ala Ala Ile Gly Gln Gly Gln Ala  
 115 120 125  
 Thr Gly Asn Ala Gly His Ile Ala Gly His Leu Glu Thr Val Leu His  
 130 135 140  
 Gln Pro Glu Glu Leu Asn Thr Arg Arg Thr Cys Asn Ser Cys Glu Gln  
 145 150 155 160  
 Leu

# (2) INFORMATION FOR SEQ ID NO:296:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (E) MOLECULE TYPE: protein

## (F) SEQUENCE DESCRIPTION: SEQ ID NO:296:

His Ala Arg Glu Tyr Glu Pro Gly Gln Pro Gly Met Tyr Glu Leu Glu  
 1 5 10 15  
 Phe Pro Ala Pro Gln Leu Ser Ser Ser Asp Gly Arg Gly Pro Val Leu  
 20 25 30  
 Thr His Ala Leu Glu His Thr Ser Asp Ala Gly His Ala Ile Arg Leu  
 35 40 45  
 Ala Ala Ala His Leu Lys Ala Ala Leu Asp Thr His Leu Val Ala Ser  
 50 55 60  
 Phe Ala Ile Asp Glu Leu Leu Asp Thr Arg Ser Arg Arg Pro Leu Met  
 65 70 75 80  
 Thr Phe Lys Thr Asp His Phe Thr His Ser Asp Asp Pro Glu Leu Ser  
 85 90 95  
 Leu Tyr Ala Leu Arg Asp Ser Ile Gly Thr Phe Phe Leu Leu Ala  
 100 105 110 115  
 Leu Glu Glu Pro His Leu Thr Thr Glu Arg Phe Thr Thr Ala Val Arg  
 120 125 130 135 140 145 150 155 160 165 170 175

145                      150                      155                      160  
 Phe Gln Gln Pro Gly Ala Ile Ser Asp Phe Gln Pro Phe Asp Leu  
                          165                      170                      175

## (2) INFORMATION FOR SEQ ID NO:297:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 178 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

Lys Pro Val Lys Glu Pro Val Pro Ala Leu Pro Pro Val Pro Pro Thr  
 1                      5                      10                      15  
 Pro Ala Leu Pro Pro Leu Pro Pro Leu Pro Pro Val Pro Gly Phe Pro  
                          20                      25                      30  
 Thr Val Pro Pro Pro Gly Ser Met Ala Pro Leu Phe Arg Pro Phe Ser  
                          35                      40                      45  
 Pro Ala Pro Pro Ser Pro Ala Leu Pro Pro Ser Pro Pro Leu Pro Pro  
                          50                      55                      60  
 Leu Val Gly Val Ala Ala Trp Leu Thr Tyr Cys Ser Thr Gly Pro Ala  
 65                      70                      75                      80  
 Leu Asp Pro Leu Ala Val Ser Ile Ala Ala Ser Met Asp Pro Pro Thr  
                          85                      90                      95  
 Thr Thr Cys Glu Ala Ser Pro Ala Ala Ala Ala Ala Gln Leu Cys Arg  
                          100                      105                      110  
 Gly Ser Cys Asp Leu Ala Pro Ala Asp Glu Met Met Gly Thr Thr Gly  
                          115                      120                      125  
 Ala Cys Gly Arg Leu Gly Glu Ala Ser Ala Gly Ser Arg Ser Arg His  
                          130                      135                      140  
 Thr Arg Arg Cys Ala Ala Ala Ser Glu Ile Cys Arg Leu Arg Cys Thr  
 145                      150                      155                      160  
 Arg Ser Ser Ser Gly Val Pro Arg Asp Trp Val Ser Pro Leu Ala Pro  
                          165                      170                      175  
 Pro Leu

## (2) INFORMATION FOR SEQ ID NO:298:

## (i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 221 base pairs
- B TYPE: nucleic acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

## (ii) MOLECULE TYPE: other: tRNA

```

ATTTCACCCC  ANGCAGCTAC  CACACGGGGA  CTCGGGAAACA  CCGGCGATTT  TACACCGGCS  180
CCTTCATCTC  CGGCAGCTAC  AGCAACGGGT  TTTGTGGAGT  GGAAATTATC  AGGGCTCATT  240
GGNTGCACCC  GGSCTTRCGA  ATCCCTCGKG  CCAATTCAAC  TCCTCNACAA  GCTTGCGGCC  300
GCATCSAGC  CCGGGTGAAT  GATTGAGTTT  AACCCTNAN  CAATAACTAG  CATAACCCCT  360
TKGGGCTCT  AAACGGGTCT  TGAAGGGTTT  TTTGCTGAAA  GGANGAACTA  TATCCGGATA  420
ACTGGCGTAN  TACGAAAAGC  CGCACCGATC  GCCTTCCCAA  CAGTTGCGCA  CCKGAATGGC  480
AATGGACNC  CCTKTACCG  GSCATTAACN  CGGGGGTGTN  GGKGTACCC  CCACGTNACC  540
GCTACCTTGC  CANNSSCCTN  RSGCCGTCTT  TCGTTTCTTC  CTTCTTCTC  CCMCTTCGCC  600
GGTTCCCTTC  AGCTCTAAAT  CGGGGNNCCC  TTTMGGGTTT  CAATTATTGC  TTACNGSCCC  660
CCACCCCAAA  AAYTNATNG  GGTAAATGTC  CCTTMTTGGG  CNTCCCCCTA  WTNANNGTTT  720
TCCCCCTTNA  CTTTGRSTCC  CTTCTTATW  NTGAMNCTNT  TTCCACYGGA  AAAMNCTCCA  780
CCNTTYSSTS  TTTCTTTTGA  WTTATMRGGR  AATTSCAATV  CCGCVTTKGG  TTMAANTTAA  840
CYTATTTTNA  ATTTTCCCGM  TTTTMMNATR  TTNSNCKCGM  FNCTCCNRKA  SSGNTTTCCT  900
CCCCCYTTSS  GKTVCCTCRN  G

```

921

## (2) INFORMATION FOR SEQ ID NO:299:

## 1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1082 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 11) MOLECULE TYPE: Genomic DNA

## X1) SEQUENCE DESCRIPTION: SEQ ID NO:299:

```

AATTCGGCAC  GAGATANGGG  CCGACCGGGG  TCGGCAAGCG  CCGGACCCCT  CCGCAGCATC  60
ACCGGGGTCA  ACAGCACCAC  GGTGGCGTCC  ANGCAAGAGG  CCGCCTGTAT  GBCGGCCGAG  120
ACGGGCAACA  CCGCCCGTAG  CAGTCGGTTC  GACTCGCGCG  TCGCTCGANC  CATGGCCGCG  180
CUGGCTGCTT  CGAACANGCC  TCGTCTCTCC  ACAGTTTAGC  CAGCANCCAA  ACCGCACCCA  240
GAAACCCACA  CCGCCCGCCG  CCGCGANACC  TCGGCTATCG  KCTGCTGGGG  CGANATCCCG  300
CGATCCCTNA  CAGATGAGCC  CTTGCCCGAA  CCGCGCCGCT  CCGTCCGGGG  AGCCGGCTGG  360
CGCGGGCAAC  CGCGAAGCCA  CGAACACGGC  AAGCACTATC  ANCGCAACAG  CAATTGTCAA  420
CGGCTAAAGC  CTTGACATCC  AGGATCTCTC  CCGCGCCACA  CGCTCGGTC  TGCAGSGCSA  480
TCCCTTCCTN  CGCGGCGCAC  TCTCAAAAGA  TCGNGATCNA  CAGKCTAGGT  CTTCCGGCCGA  540
TATGSAAGGN  CGCAACGGNT  TTAAAGCGGC  GAAAAAATC  TCCANTGGA  TAAAATCAGC  600
CGCGGANCCT  CGCGTGSCMM  NGTCVCGGKC  ATTCTTCAAC  MGGTTTNACC  CGGKTCGNG  660
CGTAACTKGC  TAAAMTTAAG  CTNCGGNTY  CCGCGCGGTA  ACCGGCCTTK  NGCCCTTAA  720
AAAACCGGNC  CTTCTCTGAT  TMMACCGCN  TCGGANTTB  CGGCTGCTTC  CANGNTYAA  780
AMGCTCTCTC  CNGGCTTGGG  GAAACCTTTC  TCGGCGCTT  CTCTCTCTCT  CMMGCGCGCG  840
AACTCTCTTC  TCTCTCTCTT  TAAAGGCGAT  TAAAGGCGAT  TAAAGGCGAT  TAAAGGCGAT  900
TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  960
TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  1020
TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  TCTCTCTCTT  1080

```

1082

## 2) INFORMATION FOR SEQ ID NO:300:

## 1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1082 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 11 MOLECULE TYPE: Genomic DNA

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:300:

```

AATTGGCAGC AGTGATCGCG CTGAAGCCCG TAGCGCGGGT GGCTCGGGTG GTTTGCCAAC      60
RAAATCCGCT CGANGTGGTC TCGGTAGCGG GTGTCCANAA CGGTGGCGCG GTGCCGGCGG      120
ATCTGATCGG CGCGGCCGTA GTGCACGTCG GCGGGCGTGT GCAGTCCGAT GCCGGAATGC      180
TTGTGTTCTG GGTGTACCA GCCGAAGAAC CGGTGCGAGT GCACCCGGGC CGCCTCGATC      240
GACTCGAACC GTTTCGGGAA ATCGGGCCGG TACTTGAAGG TCTYGAAGT GGCCTCAGAC      300
AACGGGTGTG CTTGCTGGTG TGCGGGCGTG AGTGCAGCTT GGTGACACCG AAGTCGGCCA      360
NCANCAATGC CACCGGTTTG GAACTCATCC ACAACCCCGG TCCGCGTCMA GGTCACTTGT      420
NCGGCGCTAA TTNYTGGGC GGCAAGGTTT TCCCGAYCAN KCCGCTCGGC CAAAACCTCG      480
ANTCNCSCCA AGGCCNCCAT CCNCCCAAAC AMGTTACGGG ANAAAAATY CAAAGAYCAC      540
CYTCCGGKTN TTATANCTYC CCYTTTGSTY GCGCCCCCGN CYVTGKQKAT ACCCCTNCCA      600
AWTCCCAACN CCKKCCAANA RCYKGGGGCC CCNCCAACC CGGKGGAACA WTAATTTAAA      660
CCCYAACMAW ACTWMMNACC CNNGGGSCCY AAMCGTYNR AGGTTTTSTC NAAAGAAASA      720
ANTCGGAAMC CGGNTSTACC AAAAASCCCK CCNWTCCCTC CRADATTGSC NCESAANKSA      780
AKGCCCCUNY TCGGCMWNNC CCGCGGKKKT KKGTTNCCCT WMPCWMWYTS GGCNASCNN      840
CKYSSMYCC CCGCTCCCCM CTCGCKTCC CCAMCCYANC MGGCCCCYTM GKKCCCKNT      900
YKCCCCCCCC AMNNNNGGGG WGACCCCTNGG CCGCKKRRGM TCCCNANTGA MCCTCWGNRA      960
MECYCCNRAR ANMCCSCNCC NGCNCCKKN

```

## (2) INFORMATION FOR SEQ ID NO:301:

## A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 223 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## 11 MOLECULE TYPE: Genomic DNA

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:301:

```

AATTGGGTTG GCAACGCGGG CCTGTTCCGC AACGGCGGGG CCGGTGGTGC CGGTGGGGCT      60
ATGAGTGGCG CGGCGGGCGG GGGCGGTAC CCGGGGTGGT TTGGTCATGG CGGCGCTGGC      120
AGGTGGGTTG GTGTANGTCC CGCGCGGGCC AACGGTGCTA CGCGCGGTCA GGATCGGGCC      180
ATGAGTGTTC CGGAGTCCGA CACACCTGCT CCGGCTCGTG CGG

```

## A. INFORMATION FOR SEQ ID NO:302:

## A. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 415 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## 11 MOLECULE TYPE: Genomic DNA

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:302:

```

TAGGACAGGC GTTGGCCGCA GTCAACGGCG CCATCCAGGC CGTGACCGGG CCCCCCTGAT 300
GGGAACGGG CCAACGGGG CCGGGGCAAC GGGGCCCCCG GCRGGCACGG CCGGTGGTTG 360
TTGGGGGGG GAAGGAACGG CCGGTCCGGG GTCANCRGG GGGCGGGGGG AAATGGCG 420

```

## (2) INFORMATION FOR SEQ ID NO:303:

- (1) SEQUENCE CHARACTERISTICS  
 (A) LENGTH: 1049 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:303:

```

AATTCGGCAC GAGGGGSCAG ATCCCATACA TCGCTCCCGG CAGACCCGGT GATACAGCA 60
GTTGGGACCA GCGGAGCGCA CAATACGGCG TCTGGCTGTC CCGCTTGAGC ACCACCCCT 120
TACCGGCCAC CAGCGCGGGG ACCGATCGG ACACCGTAAG CGTCATGGGG TAGTTCCACG 180
GGGAGATCAC CCCCACCAAG CCGTTGGTT GATAGCACAG CGTGGTCTTG CCTATCCCG 240
GCACGAGCGG CTGTGCCTTA CCGGGTTTCA GCAGGTCCAG ACAGACTCGT GCSTTATAAT 300
TTCGGCTTCC GCGATCAGAT CGACAATTG CTCTTCCCGG GCGCATCGGG CCGTGGCCCG 360
CTCGGCTTGC AGGAAGTCCA TGAAGAAGTC GCGGTTCTCG ATNAACAGGT CGGATAGCG 420
GCGGATCACT GCAGGTGGCT CGATNACGGG ACCTTCGCGA CTCGGTCTGC GCGCGCGGAN 480
CTTCGGCGAA TGCTGCTTCC ACTTCGGCGG NCGTGCCAA CCGAATCNTAT CAGGGGTTGC 540
CGGTTAAAC TCTCAATST NCGGTGAA ATTCGGCAAC TTCTTATCCC GCGAGGTRCC 600
AAGSANNCAA ACCTCGGCAA GGTAGGMIT TCGCCCNCTT YCAAAAATNC GGTCTTTGGN 660
CGAATTTCCG CKCNATGKTG MCAAGGTTCT CKAANAACGS GGGTCYTCTN NTGNGKGGAK 720
CGAAMGGHT TTGGGMAAG CKNMCCAAH GTWAGGCTG KTKAANGGNW TTCCCCCGGG 780
GGAKKNGA ATYCCGNA NCCCRGGGG GNMCAATTC TYCCGMCCTC CTCGGGAWTC 840
GCGGCTTCC CAAAAAACG CCGAATTMM TCTTTGRCN TCTGANAAG CTTTARGCA 900
MMSSAARNG NMNCTGCTG CHTTGTGKT AAAAAGNAY CCGMAAATT TTTAWTTCC 960
GCGCGGGG NCGCTNTTT TCGNMTWTH WNTNORMCG MMSNCKNSG HKGNPCNN 1020
TCTGCGCG AAWNTKGYN TTTATMAGC 1040

```

## (2) INFORMATION FOR SEQ ID NO:304:

- (1) SEQUENCE CHARACTERISTICS  
 (A) LENGTH: 100 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:304:

```

AATTCGGCAT GAGGGAATTC AGAATCCCGG AATGCTAAG CTTGGTCTTC TCGGTTACG 60
CTAAATKTCG GGTGAGTCTT TCGGCTGTC CAAATCTGA GGTAAATCTG AAAAGCTG 120
ATGCTGCGG TGAATTAAT TAACTTANT TAAATGAA AATGCTATT TAAATCTAT 180
TCTGATCA AATCTCTAT TAACTGAA TCTGATCTT TAACTGAA TCTGATCTT 240

```

YTTATTTTMS	GCTNAYGGGA	ATBAMRGGAA	CAAYNTCCCT	CCCMGGAAAA	ACCAACMSGC	540
CCTGGTNSYC	CNCCCRCCNC	AKAACCCRTT	KCTGTRSTMC	CCSMAAATNA	CSCCCSCTTS	600
NACTCCNCSC	AANTNSCCCC	CCCSCKNNTT	ATSTYCCCGK	GTTCCCGGMC	CCCTTNAAMC	660
TCCCCGGTTA	ACCCGCTWNT	SNCNCCCCCS	YTAAKMNCRG	GCTTSTTNGT	CCCCCYTRMK	720
CNCCCCCTCK	SAMCWNCNC	CTCKAACNAC	CCCKCYKGS	TNCCCAATNT	WCMWCKCCNS	780
KTNTMTCTKC	CCAAYTNCRC	CCNCRCTCCC	CKSTSTCAM	WTATAAAACG	WCWYAWYNK	840
KCNWMAWTA	MGACWCTCNY	NCCCNCCNCK	NTTKTAMWCC	CKMCCCKCSW	TWCYCKCSCC	900
CCNTCTMNAC	YCCCCCKKTY	NKWMCCCTTC	CCCCCTCCC	MCNMBMKTCT	YCSGKTWCWC	960
NCYNTTMTCN	CYNANMCKCK	KTCTCTTCN	CRNTCTCCCC	CCWCCCCCCT	KKCTCTSKCC	1020
CNCNCTCCSC	MMKGS					1036

## (2) INFORMATION FOR SEQ ID NO:305:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1036 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (B) MOLECULE TYPE: Genomic DNA

## (C) SEQUENCE DESCRIPTION: SEQ ID NO:305:

AATTCGGCAC	GAGATCATCA	ATAGCGGGCT	GGTCAGCACC	GAAGTGGTCC	GCGATCTCCG	60
TAGCAATCT	CGTCTGCTCG	CCCAGCAGGA	GGTCGGCACC	GATCGCGACA	CCTGCCGATG	120
TTTGGATGGT	GTTCAGTTCC	AGGTAAGGCC	GAAGCGGCAG	CTTTGCTAGC	AGGGTGTCTT	180
TGCTTTTCGC	ACGTGACGTA	ACCAATAACT	GGAGCGCAGA	CCAACTCCGG	CCCTCGATCC	240
GGGTACGAGG	GTCCGCGCGA	CCCAGCCGTT	GTCCGCGCTG	GGCGGAAGGT	CAGCTGCTGT	300
GCGATCGAAG	TAAGAAACCG	CGCATGCCC	GTGGCAGAT	ACGACTGACC	GAGCAAACGA	360
ACGATCGTCC	TGCTTTCCCT	GGGGGTAAT	GAGCCAGCA	ACCGCAGCAG	CCACCAATCA	420
TTGGGATTCG	CGCACTGACC	GACCAACCGT	GTCTGCGACA	CCCCAGCGGA	ATTGGTGGTC	480
TTGGCGGGGG	CGGCAACCGG	AATCAGCCGG	ACCGGCTCCG	CGAASCACCG	GCATAGCCCT	540
ACATAGCAAC	CGGCTGCTCC	CGCAGTTTC	GGGTTTCTCC	CGCTGCGCAA	CGGCAAYNCC	600
TCGAATTCG	ACGCAAAAAA	TTGCGCATY	ARNGTCTTC	CGAAGAACCN	ATTCGCTCTA	660
TGCGCGGGGG	GGGTCGCGCT	NMNAACCGG	CGGCAAGCCG	CGGCGCGCCG	CGGCTTCTCT	720
CGCTTCTCT	CGGCGCGCGG	TTTGGTCTCM	GGGCTTCTCT	CGGCTTCTCT	CGGCTTCTCT	780
AAAAAYCKG	CAAAATYAAA	CCGCTYMAAA	AKCTCGGSSC	CGGCTTCTCT	CGGCTTCTCT	840
ANTTAAGCCN	CAAAAAAAWW	NCANMCCCG	NGGCGCTTAA	GGCTTCTCT	CGGCTTCTCT	900
AAAAAATMTC	TANATMNSSY	TTNAAAAAAA	AGGCTTCTCT	CGGCTTCTCT	CGGCTTCTCT	960
CGCTTCTCT	TTNAAAAAAA	AGGCTTCTCT	CGGCTTCTCT	CGGCTTCTCT	CGGCTTCTCT	1020
CGCTTCTCT	TTNAAAAAAA	AGGCTTCTCT	CGGCTTCTCT	CGGCTTCTCT	CGGCTTCTCT	1036

## (2) INFORMATION FOR SEQ ID NO:306:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1036 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (B) MOLECULE TYPE: Genomic DNA

```

GCCATGGGCA ACGGCTACTG GGGCAACCCG AATCCATTGG GCGTCTCACC GCAACCCCGG 120
AAACCGGGCA CCGCGGCGATG GATCAACCCG CCCACCCGAG ATCCGAAATA GCCTCCACAT 180
AATGAGACAC TGGCGCAAAG AGCTTGACAG GCGCCGCACC ACGCAAGCTG TTAGACGTGT 240
CGGTCTTGCA AGAAGCGGGT TGGCCACCCA AGATCACGCC GCCCAAGGGC ATCGAGTCAA 300
CGTTGCGGTG GTAACGCGCT AACGTGGGGG CCGCCAAGAA ATGACGGTGC GCATTACCAT 360
GGCCCTGCTG ATACCTTTTG GCCACCTGCG CACCANAACAT ATGANCAGCC TTATGCCGAG 420
TCTCGTGGAC ATCGGCAGCC GCTTCAAAAA CTCCTTGTGG ACAATSGTAT TGCTGANCCG 480
CCGAATTCIT NTRCTTGCAA SAACACTNCA TGTTCNSGGT NAACAACCTT GGTTCGAAAA 540
ACANCCAATA TTGAANTCCC ANTGGGGCAG GAACCNCTTM CGGAAGKTGK TGGGAACGAA 600
TGKTGCCCCA AAATCCCGGG NGGTRAAAWW CCCNSNATGG MSAATTTTSC CTNGAACAAAM 660
AAAAGGTCCA AGKYCAAAGG NGCCCCCCCC SGNAATTTGG TGAACSCAKA WYANRTTCCC 720
WWWTCNAAAT MTTNGGGTCC KNNTCCCGWT AAANGGGSCN CCCNCNCRGG GMGTYTCCCC 780
NWNMGGMGN CYVSCCCCA AAAAAAAMMM MTTTCSGKGG SMGGKKCCCC CCSGGTYWGG 840
GKKYTTAAAC CCGGKGGGTN CAAAAAANAN ACCCCCCAMS NGGGGGGAAA ATTTGNAAWT 900
AAGGKKKTKC SCMACCCCAA AAANMMNNCN AWCNCCGMGK SARGGGGRNY TTMKAGGGMG 960
GNYCCCCCW YCGGGGGGNA NAAYAAAAGK NGSNGRGAAT NTTTTTTTGT RSSSRNKTTT 1020
TYNTCTTCN CCMGNRWWG SRAMNTGKTS NSSGGGSGCC 1080

```

2. INFORMATION FOR SEQ ID NO:307:

(a) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1040 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(b) MOLECULE TYPE: Genomic DNA

(c) SEQUENCE DESCRIPTION: SEQ ID NO:307:

```

AATTCGGCAC GAGGTTGACC AAAGAGCTGA CATGCCGGGT GATGGGACAT CCGATCGAGG 60
TCAATACGGG CATGATGAGN CCGAANGGAN TGTGGCGTTT GCTCAACTGG ATTACGGTTG 120
TCAAGGTGAA ACCTTTTGGG CCGAAAGATG CCGAGCTTAA GTTCCGCTTG CACCGTGCAA 180
TGTGTGATG GATGCTGGAA CCGCGCTGAC TGATAANGAA TTGCTGTGTC GCGCGGCACN 240
ATCGATGCTG CAGTTTTCNC TCGGCGCTTA AATTGCTGTG GCATCATCTG GCAGGCTATG 300
TTCCCGCTAC CCGCAGGCC ATCATGATG TCGCGCTAAC GAANAAGTTA TGACATGGCG 360
AAAGCGAMTC GGGCATSCNC CCGGCAMTTT CCGAACCTGC TGTGTNTGAA CCGTMTCAAC 420
CGAATCGGGC GCTYAAAAGC NGGCTTGCGT TGATTMMAAC CNAACCCNTN CNATYCTTTG 480
CCGNGMNTG CTTTCTCTCG AACTCCGKKC SYTCCGCTG TGAAGCGEMA CTNCCCCCCC 540
TTTGAGCTTA MPTTTTCAAA AAMCGOMTNA AGGCGAATNN GAACTTNCCT TCAANTAMM 600
AAANTCGGSC TTYGGNFCN CCGGCAAYN TTGCTGNGGG GNNNTYCTCN GCTTYNGGCG 660
AAACNTTTT CORTNCHMMN TTACAMCGC NCMNTMTTAM TGGGCGNNAG GWCCTGGGKY 720
TNTTTCNCAW TCHNSKTTT TNGGGGCGCG TCGYGTGTCM NTGAGCGCCC GCGGCKKMAA 780
AAAAAMCMA RCGGNGYGGG KCGGCGCCCM RNATNGGGCG TCTFAAACAA ACCCCANRA 840
TNGNMGGGC SMACCGNGN HNAAAAGCT TNSCTMANM MKGMANNNCT SGMSCCMNSN 900
NCTGCGGGCT TTKGNGAPN AANAMKMGGM CCGGNCCTN GAAAGGGGMS GCGKSCNNGN 960
NGASNGMGN BRNGANRCC NNGYENMPN NNGNNNGNNR GGGKNNACN NMKMCAWSMC 1020
NENMMGNNG CGYMTNKCC 1040

```

3. INFORMATION FOR SEQ ID NO:308:

(a) SEQUENCE CHARACTERISTICS:

(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

AATTCGGCAC	GAGACAANGG	CGTGAATGG	GATCCGGCCG	AGCTGGGGCC	CGTCGTCACC	60
GACCTGTGG	CCAAGTCGCG	GCCGCCGGTT	CCGGTCTATG	GGGCCTAGTT	ATCTGCGCCG	120
AGCGTGAAC	TTCGGGCGAG	TTTCGGCCGT	TTTCTCGCCC	TGGCTTCACG	TTCCGGCGAAG	180
TKGGGAACGG	TCAGGGTTCC	CAAACCACGA	TCCGGATCGT	GCGGTCCGGT	CAGGACTGGT	240
ANTCCTGATA	CTTKGGTACA	TCGTGACCAA	CTGTGGNCAA	TATTCGGCCG	GCTCCTCGTC	300
NGTCGCGTCC	CGCGCGGTAA	GGTCCANCAC	TTCTTTTTC	TCGTGCCG		348

(2) INFORMATION FOR SEQ ID NO:309:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

AATTCGGCAC	GAGACACCGG	GTGTTGACC	AACGGACGCT	TGGGCGCGGG	CCCTTGGCT	60
GGCATCAGCC	CTTCTCCTTC	TTAGCGGCTT	AACGGCTGCG	TGCCTGTTTG	CGGTTCTTGA	120
GACCTGCGCT	ATCCAGCGAA	CCGCGGATGA	TCTTGTAGCG	CACACCGAGG	AGGTCTTTCA	180
CCCGGCGCGG	GGGACCGAGC	ATCATCGAGT	GCTCCTGCAG	GTCTGGGCGG	TCGCGCGGAA	240
TGTACGCGCT	GACCTCGAAC	TGACTGCTGA	CTTCACCGCG	GCAACCTTCC	GAAGCGCGCA	300
TTTGGGCTTC	TTGGGAGTGG	TGGTCTGCTG	CG			336

(2) INFORMATION FOR SEQ ID NO:310:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 961 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

AATTCGGCAC	GAATTCGCTC	AGACCGATTG	AATTCGCTCG	CGACGACCTG	CGCACTGCAC	60
AGCCTGCAGC	AAAATGTTCG	GAATTCGCTG	AACGAGCGCT	TCCAGACGCT	CAGCGGCGCG	120
CGCCTGATCG	CGAAGCGGCG	GAACGGGACT	CGTGGAACTG	GGCTTGACCG	CGGCGCGCGG	180
CGTCTCTGTT	CGGCAACCGG	CGCAACCGCG	CGTGGGCGGT	GAACGGGAAC	AACGGCGGGG	240
ACGTGGGCGC	CGGCGGCGCG	GAATTCCTTC	CGACGGGCGT	ACGCGGCGCG	CGGCGGCGCG	300
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	360
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	420
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	480
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	540
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	600
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	660
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	720
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	780
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	840
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	900
CGTACAACTG	CGACGGGCGG	GAACGCGCGG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	960

NTTNCYTTKN NATTKGGNNA AAAANCCCTY CCWCSGRAC	NCCCCCNGM GRGMCNNTNN	720
NTTTYGNCNN CCGGSNAAM RNTTKATTTG NGGGGNTCN	GGGTMNNNA AACCCCAAAM	780
MNRNNKCSA ANGGGKSGC NKNMMNSGT TTTYCKMRA	MRNWTYKNKN NTCNGARSRN	840
NAAMCNSNK NGKKGNKAA ARNNTTWKTN KNSCNNCNN	GRRNGVRGGC CKMKGSNMNG	900
MCWHNAWRNG NGSNCNCKC NKNMAAAAAA AASGGVNCKS	NSMKVKKKKG NRGGGGGGGG	960
GG		962

## (2) INFORMATION FOR SEQ ID NO 311:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:311:

AATTCGGCAC	RAGAAGACGC	CCGAANGTTT	GGCTGGCTC	TACAACTTCA	TCAAGCCCA	50
GGGGGAACGC	AACCTCGGCA	AGATCTACGT	TGGCTTCCCC	GAAGGGGTCT	CGATGCGCCA	100
GTACCTCGGC	GCACCTCAGC	CCGAGCTGAC	CCAGGATCCG	GCTTGGAAAT	GGCTTGGCTT	150
GCAGAAAGATG	TGGTTGGAGG	TGGCTTGGAG	GATTTTGCAN	GCGACGCCNG	TGACCGCGAC	200
GGGTTTGGTG	TGGGCACTGC	TGGTCAACAC	CCGCGGCACC	GGTTTGACCT	CGACCAAGCTG	250
CACCACTCGT	GGGCTCTGTG	CCG				323

## (2) INFORMATION FOR SEQ ID NO:312:

## 1 SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1034 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 11 MOLECULE TYPE: Genomic DNA

## X1 SEQUENCE DESCRIPTION: SEQ ID NO:312:

AATTCGGCAGT	GTGTGTGGCG	GGCTTCAGAA	GAAGATGATC	GGGAACATCG	CCAGCGCCCG	50
TGAGGCTATG	GTGCGGCTTA	TGGGCGAGTA	AGGATCATC	AGCGGCATAC	AGCGGGGCGT	100
TCAGACCCAC	AGCAGCTTGT	GTGACCTTCA	TGGTTTGAGC	TAAGCGGTGT	AGACRAACAT	150
CTAAAGGCGC	AGGCTGACCA	GGGCGAGTA	TAGGCGGAGC	AGGTTGGTGG	GGCACCATAA	200
CTAGAAAGAA	AGATGACCG	TGACCTTCA	TAGATGCGA	AGCGCTTTGG	GGTGGGCAAG	250
TTTGGGCGC	CGAAGGCGCG	GGGCGGCTT	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	300
TGACCAAGTT	AGGCTGTGTT	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	350
CTATGAGCAJ	GGATGAATG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	400
CTAGAACTTG	GGGCGGCGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	450
AACTACATAA	GGGCGGCGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	500
AACTATTTCC	GGATAGCTGG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	550
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	600
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	650
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	700
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	750
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	800
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	850
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	900
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	950
CTGCTTTGCA	ATGGAATGCG	GGGCGGCGCG	TGCTTCATCA	CTTGTGCGAT	ATCGGCGCTG	1000

SCNSNGGKBC CCCC

1034

## (2) INFORMATION FOR SEQ ID NO:313:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 331 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

AATTCGGCAC	GAGCCACAT	CCCGGGCCGC	TGTTTGCATG	ACTCGTTTGT	CATCGTCGAC	50
RAGGCACAGT	CGCTGGAGCG	CAATGTGTTG	CTGACCGTGC	TGTCCCGGTT	GGGGACCGGT	120
TCCCGGGTGG	TGTTGACCCA	CGACATCGCC	CAGCGCGACA	ACCTGCGGGT	CGGCCGCCAC	180
TACGGGTTCG	CGCCCTGATC	GAGAACCTCA	AAGTTCATCC	GTTGTTCCGC	CACATCACTT	240
TGCTGGCGAG	TGAGCGCTCG	CCGATCGCCG	CGCTGGTCAC	GAGATGCTCG	ANGAGATCAC	300
CGGGCCCGCG	TGAGTGGCGC	TCCCGCGAGC	A			331

## (2) INFORMATION FOR SEQ ID NO:314:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1026 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

AATTCGGCAC	GAGATCTCTA	CCCTGGCGAG	CAGTGCACCC	AGCCGACGCC	ACCAGTTAGC	50
CGTGATGGGC	CAGAAGATGG	ACCAGGTGCT	GGCCATCCCG	TCCACCCGAC	TGGAGCTGAG	120
TACCGGGATC	CCCGTCTCTA	TCTACGGCGA	TRAGCTGGTG	TTCGGCATCA	CGCTGACTA	180
TGACCGCCCG	TCCGAAATGC	AGCAGCTGGT	CAACGCTATC	GAAGTGGGTG	TGGCCCGTCT	240
CGTGGCCCTC	ANCGACAATT	CGTGTCTGCT	GTTTACAAGG	ATCGCCGTAA	CGCTTCATCC	300
CGCGCACTCC	CCANCGCCCG	CGGGCGGGGG	CGGCTTCTTG	TCCCGACCCG	CGGAGCGCCT	360
TACTGACGGC	ATCTCGCTCG	CGTTAAGCTT	CTGAGAAAGG	TGGCTCGTGC	CGAAGTTGCG	420
TTCGCTCACC	ATCATCTCCG	CGCGCATCA	CGGCTGCTTG	TTCACACAGA	CTTGGACGCG	480
TCCCAAGGAA	CTGCTCCCGC	AMTNCAGGAA	NTTCTCTGCG	CGACGCGCTT	CTTGGCTTCT	540
CGCTTAAACT	TCTCATCTTT	CGCGGGGCTT	TCTGCTTTTA	CGCGGGGCGC	CTTCTTCCAA	600
ATCGGSMMAA	ATCGGCAATC	AAACCGCCCG	CGTCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	660
TAACCGCCCG	NTTAAANTCT	CTGCTGCTTG	CTGCTGCTTG	CGCGGGGCGC	CGCGGAGGCG	720
TAATCGCCCG	CGCAWTTTAA	CGGAGGCGCG	AACTTCAAGG	CTTGGCTTCT	CTTGGCTTCT	780
TAATTTGCGC	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	840
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	900
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	960
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1020
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1080
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1140
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1200
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1260
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1320
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1380
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1440
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1500
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1560
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1620
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1680
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1740
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1800
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1860
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1920
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	1980
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2040
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2100
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2160
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2220
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2280
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2340
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2400
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2460
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2520
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2580
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2640
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2700
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2760
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2820
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2880
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	2940
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	3000
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	3060
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	3120
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	3180
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	3240
CTGCTGCTTG	CGGCAANTAA	ATTCCGCGCG	CGCTTCTTG	CGCGGGGCGC	CGCGGAGGCG	3300

- (A) LENGTH: 324 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

AATTCGGCAC	GAGAAGACGC	CCGARNGTST	GCCTGGCTC	TACAACTTCA	TCAARGCGCA	60
NGGGGAACGC	AACCTTCGGCA	AGATCTACGT	TCGCTTCCCC	GAAGCGGTCT	CGATGCGCCA	120
GTACCTCGGC	GCACCGCACG	GCGAGCTGAC	CCAGGATCCG	GCCGCGAAAC	GGCTTGCGTT	180
GCAGAAGATG	TCGTTTCGAGG	TGGCTTGGAN	GATTTTGCAN	GCGACGCCNG	TNACCGCGAC	240
GGTTTGTGTG	TCCGCACTGC	TGCTCACCAC	CCGCGGCACC	GCCTTGACGC	TCCACCAGCT	300
GCACCACTCG	TGCGCTCGGT	GGCG				324

(ii) INFORMATION FOR SEQ ID NO:316:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

AATTCGGCAC	GANGCSTGCC	GGTNAACACC	AGCGCGCGGC	TGCCAGATAT	CGCGGACTCG	60
GTAGTGGCGC	CGGTGGCGCTC	GTGCTCTCC	TGACGCGGCG	CGGCGACCAT	AAGGTGCGCTM	120
ATGCGCGAGT	AGCGCGCGCG	GTGATCGAG	TGCATGATGA	TGCGACTCTC	CAGCTCGCGC	180
ACCGCGAGCT	TGGCATCGCG	GTGATCGAG	CAGGACCGCT	AGGACAAGTC	GATCGAATGC	240
ATACTGGCGT	CGAGAGTGGC	GTGCGAMTTC	TNGCTGCTC	CACCGCAAAT	GGCTTGATTG	300
GTACTGGCGC	TANTCTTGGC	GCATCGCGTC	CGGATCGAAT	GGGAACCGCA	GGATGGCGAC	360
GAACCGGTCT	GANTCGAGCT	TTGCGCGCTT	CGGACAGTC	GTGACANCC	GGTACTCGGC	420
ATAATCTGCG	CGGNAATCG	CGTGGGACCG	TGCGGACNAT	AANAACGGGC	ACNACAATCG	480
CGCGCGCGGT	CAGCGNAACA	ACANCTTGGC	ATCGGATTTT	GTCCCGCANG	GTCAANCGCT	540
CGGGAACCGC	TGCTCGCGCG	NACCTTCTCT	NNAWTAATCG	CGGCTTCCGH	CGCTGGNGCA	600
ATAAATCGGA	AACCGTTNCC	CGAGCTTGAA	GGGCTTCTTC	NATTTTACT	GSTAACCGCG	660
AATTTTTCGG	GANTCGCTCG	CGCGCGCTTC	GTGTTTCCCG	ACCTTNGNAN	GGCGCGCGCA	720
AGTTTCTCTT	GTGAAAGCG	GAAGCGGAA	GTGTTTCTTC	NACCGCMNAA	MYMTTTCGG	780
GAAGTCTCTT	CGCTTTTAA	CGCGCGGCTT	GAAGTCTCTT	CGCGCGGCTT	CGCGCGGCTT	840
GAAGTCTCTT	CGCTTTTAA	CGCGCGGCTT	GAAGTCTCTT	CGCGCGGCTT	CGCGCGGCTT	900
GAAGTCTCTT	CGCTTTTAA	CGCGCGGCTT	GAAGTCTCTT	CGCGCGGCTT	CGCGCGGCTT	960
GAAGTCTCTT	CGCTTTTAA	CGCGCGGCTT	GAAGTCTCTT	CGCGCGGCTT	CGCGCGGCTT	1020

(ii) INFORMATION FOR SEQ ID NO:317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:317:

AATTCGGCAC	GANGCGTGCC	GCNAACACC	AGCCCGCGGC	TGCCAGATAT	CCCGGACTCG	60
GTAGTGGCGC	CGGTGGCGTC	GTGCTCTCC	TGACGGGGCG	CGGCGACCAT	AAGGTCCCTM	120
ATGCCCAGGT	AGCGGCCAG	GTGCATGGAG	TGCATGATGA	TGCGACTCTC	CAGTCCCGC	180
ACCGGGAGCT	TGGCATCGGC	CCTGATCAGC	CAGGACGCGT	AGGACAAGTC	GATCGAATGC	240
ATAGTGGCCT	CCAGAGTGCC	CGTGCAATTC	CNGCGTGCTC	CACGGCAAAT	GCCTTGAATT	300
CTACTCCGCG	TANTGTTCCC	GCATCGCCTG	CGGGATGAAT	GGGAACCGCA	SGATGGCGAC	360
GAACGGGTCT	GANCTCAGGT	TTGCCGCTTT	GCGCACAGTG	GTGNACANCC	GGTACTCGGC	420
ATANATCTGG	CCCNAAATCG	CGCCCGACGG	CGCCACNAT	AANAACGGGC	ACNACAATCG	480
CCGCCCCGGT	CACCCNAACA	ACANCTTGSC	ATCGGATTTT	GTCCCCANCG	CTCAANCCGT	540
CCCGAACGCC	TCNTCCGGCG	NACTTTTCTT	NNAWTAACTG	CCGCTTCCGK	CCCTGGNGCA	600
WTAAATGGGA	AACCCCTTNC	CCACCTTGAA	GGGGTTGTTG	NATTTTACT	GSTAACCCCG	660
AATTNTTCCG	GANTCGGTEN	KCCGGGTTTT	YSTNTTCCCC	ACCTTNGNAN	GGGCCGGGCA	720
AGSTTTTCTT	SYTGAAGGGG	GAAACCCAAC	TTTNTYTYTY	AACCCCMNAA	MYMTTTYTSG	780
MNAASCNK	CCCTTTTAA	CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCGCCYMANG	GGGGGFAAAA	TCTKTCNNCC	GGGCCKAAAW	ACCCMMMYGN	TTGKFKYKSS	900
CCSAAATTTT	NMMRAACTKN	GGGGCCSSGA	NNTTTNAAAG	MSCCCCCSNN	STGFCCTNN	960
NTTCCCNAA	NMKKGHNWNM	BNMNSCSNGG	CKYNSGGGNN	NNAAGMGGGG		1010

## (2) INFORMATION FOR SEQ ID NO:318:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1092 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (2) MOLECULE TYPE: Genomic DNA

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:318:

WONGGGGWS	NTCAVCAYCA	YCACGGGGYW	WATTCCGGG	CGCAWCTTGT	MAASAGATGT	60
CGAAYTCGGC	AMGAGGGAMT	CKCTMGCCNC	CTGTGGCAAN	CGAATRAGGC	CTRATAATTT	120
CGACTCCACA	AAAAAGCTTT	CTGTGTATYT	ACCGRAAATR	AAGGCGCCCG	TNTCAACWYC	180
CCCGGKTTY	CCRATYCCCG	TKTTGTAMCT	GCCKGGGTSR	AAAYCCCCCG	TGTTGGAYCC	240
CCGGATTGAA	ACTGCCCGKT	TCAAACTGCC	TKTTTSGCSA	CCCGGKWATT	GAMSTCRCCG	300
ATTAAAAAAC	CCGKTTTGGN	CTGSGNCCTG	CGAAATNCGE	AYCCRATAYC	CCATGCCCTG	360
ACTCTCTCC	CCGTACCCA	AACTGGGTA	TCCTATACTG	CTCCCTAAAK	CTAAWYCKGG	420
CTCTYIMMTY	TTGCKGGGCT	CTNAATTTAG	CAAAATCTCT	CTCTTCCATA	CTNAAGCNCG	480
CTCTGUCWEE	ACMCNCRAAA	AAAKAATAAT	PAKAAGGTTT	CTATNYGAAA	ACGNCGGGCTN	540
CTONANTNEN	ATCCONTNCC	ACGNCGCCCA	CGGCTNAAGT	CTSSGAAATTT	CTMMAAGCCG	600
CAANCCGCA	CAACNTNCCG	CAABAAAACC	CTYCNCGGGG	CTCNWNCAAA	ACASONTTAT	660
CTCTKSTTT	CTGGMWCCCT	CTCTGCNAAA	ACCGAAASTA	CTTTTYTGGGT	CCNAGAKAAA	720
CTYNGGGGYN	CTMCCCNAA	CTATATCTCT	CTGCAANCCT	CTAAACCTTT	CTMNAACCNK	780
CTPMTCTCTT	CTGCTGCAAT	CTGCTGGRAT	CTGSGNCCTY	CTAAAKKKK	CTAKWWNNGNG	840
CTPNNACGMA	ACCGGAATY	CTMNAAAATN	CTCCCCCTCT	CTNAACACGNK	CTVTCGSAAA	900
ACCTGNCCTT	CTGCGGURAA	ACCGGCGNA	CTANTNCCCA	AAAACTNYNGK	CTTCCGCGCC	960
CTAACMAAAA	AMCCGCGGSM	CTAACCTGCTN	CTMCCCNKYY	CTTTTCTCTT	CTTCTMPCCTG	1020
CTATCTAMWGT	CTSKINMAAAA	CTAACCTANTN	CTTCTCAAACT	CTCTNYWPSW	CTGCTMCMNA	1080
CTATCTMCTT						1140

## 1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1251 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

## 11. MOLECULE TYPE: Genomic DNA

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:319:

```

GGGGGGGNNN NATACATCWT CYGTGYACCG GGGMTCTAKT GGGGGGGGGG AATCTNGTCA      60
ASAGATCTCT NAMTTCGGGC ACAAAAACWT GACAAASYMT CGNGCNMTCC GTGTCCCTNKA      120
TCGCAAAACG NGTRACASAC ASACACRTAT GTGTGCCCCAC CASCAYTCK FTGGGACCTC      180
GCTRACCGGY TGGGCRNACG CCACGYTCCG CWTCTATCCC RACGCGGGCC ACGGGYGGGG      240
ATATTCCAGG CACCACGCGC AGTTTGGTGG ACAATGCGCT GGCATTTTCC TCRAANTTCC      300
TGAAACCGAA TTGNSMTTGA ACCNCCAARG CCGCSNCCNR AACARTTGGG WTCCGCGGTT      360
CTGCGGACCG KTTTCGGGGG GTNTCGGCAN AANCGCACCC WTGGWTTCTM TCNCCGCAAC      420
GGGGGACAA NTGGGGTTCG AATTTTCCRA AYCGGGGCGG GGATTCCSCA AACGGGTGCG      480
GAAACTGTTT GCRAMACCG GGAACCGCAA TTTCCGGGCR ANAAATTTCTN YCNCAACCACT      540
GCTTCTACTT CCGCGACCGT AACMANTTTC ATCGTCTNNN CCGTGTGCGCT TGGGGCAGGG      600
CKAAAYACCG CMTTKGGITT CGCAACCTGC GGGCCCAANTC CCHAMCCRCA CTTTCNATTT      660
GNTGCAATT GCGCGCGGCT RANAACCGCC NTGGCCNRYT CGGASSAAAA NCGGCGCTNT      720
KGGCNSCCCC AGTAANACCG TACCNNAVTS CAWCTCTTGC CAAASTTKGG ACGAANSKTG      780
GGNTTCCGGK ATTTYTTTGS GGNCCGCTN TATNGGNTN GGGCCKCYNC NCSTKTGKCA      840
NASSKAYCCG NGNKGGGGGT ACCCGGCTMG GGGGTTTTTT NSSGCCCCCC AWAYGNKSTG      900
GCCCCGNNCG GGAAKAATWT MWTMCNSGG GGGAAWTTTT NTSTGGAMCS SGGACYCCCP      960
GGGGGKTTTT TCCCCNCBA MNAWANGGGG GGGGGANAYT NTGNSGNGGG KWNTTTTATTT      1020
YTYCYCCTM TKACMSGGGG GTTTKKAENG GGGGGAGAAA ANAAAAAARA RAKGGYKNTT      1080
TGNACACNTT GKNWNWNR NAGAGTCTT CKCKCCNCGG SNTTCTTTT MGNSSGYGGG      1140
NRGNRNAAA ACNKRMMAC KCSYTYCCG GSYCTCCTCC NCNCGGGYGS NGSCONSTYN      1200
NNKGRKNTA TMTMGNCGT SCCTCCNCGG SKNKTCTG TMTCNMYGGG C      1251

```

## 2. INFORMATION FOR SEQ ID NO:320:

## 1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1099 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

## 11. MOLECULE TYPE: Genomic DNA

## 1. SEQUENCE DESCRIPTION: SEQ ID NO:320:

```

GATTCGGAAT GAGATATCAG CAATCTTCTT GGGGTAATAA AATGGAGGTA TTAATATCTT      60
TATTTGATTT TTPACATCTT TCCCGGCTAG KTGCTGGGTT GATGGGTGGC CTGGCKTGAK      120
CTRAAGGTTT TGGGCRNACG CCGATCGGCC CAAAGCGTTT TGGGAGGACG ATCAKCGCGG      180
GAAATTTTGA TTGMCAGCGG ACCCGGGGCT GNCATAACT TAAAAAGGCT GGCMCTGCTG      240
CTGGGCTNAT CUNTSTUGCA ANTGGAACTC AGCGGCGNMA TATKMAAGTA MAAGCGTTT      300
CTTAAAGCTA TTGAAAAAAT CTCAATWTTT GCGGCTACTT TATGAAAATG TNGAGTGYA      360
CTGAGCGGCT CCGGCTTCTM TCGAAGGCTT CTTTGGGNTG GTTCAMCTTN AAAMCAAGTA      420
TTGCTTCTCT TGGGTAATAA GGGGTAATAA TTTTCTTCTT TTTTCTTCTT TTTTCTTCTT

```

(A) LENGTH: 296 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

AGAGTATACA	TGCTGTGYA	CCSAGGATCW	ANTGCGCGCG	MAAKCTWSTM	CASAGATCTC	60
AAAYTCTGCA	MGAGCGGCAC	AKAKYSTCGT	CGMRACCGCG	CAACACCCWC	CNCGCCCCWT	120
CTTRGACCGG	GGCKATASMC	ACGTTTGGCG	CGCGGCGCA	CGTACACCAC	CCACGCGCGC	180
AGCGCGCGCW	TRAMCAAACC	ACCGCGCKTT	TACCGCGCGC	CGCGCGCGGG	CCACCGCCAG	240
CGCGACCGGC	ACCACCGCGG	CGCGCGTTGC	CAAAACAGGC	CGCGCKTTTGC	CACCGA	295

A: LENGTH: 1023 base pairs  
B: TYPE: nucleic acid  
C: STRANDEDNESS: single  
D: TOPOLOGY: linear

YNGGSGNKMY	ATCATATWTH	YNGGSGN IG	MTGNAATTGG	AGGTAATTT	TGTMNASAGA	1
TGTGAAAYTG	GGGAMGARGA	TGTGCGGCGN	TAATGTCCAA	AGGTGKHTAA	GGGIMATGGG	10
YTGCGGCGYCA	AGGAGGTET	TAATATGCCC	GGCAMWTYCA	AACGATTAT	YTGCGYCGAG	100
AAATATTTCG	CGTGTGASCA	AGTGTGAGCG	YGTGAASCA	AGCGGTCTTA	AGGTAAATY	100
YTAGGTTNKT	YGGGCGAACA	AGGTGATATA	TGCGGCGCGG	AGCGACAAAA	TGTGANTNGT	300
YNTGTCGAA	GGGCTTCCG	AGGGGCTTA	ACTGCGGTAR	CGTGTGCGY	GGGTTCACAT	360
TAAACCGCGG	CGGNTGWTG	GGGCGCGCA	ATGCGTGGC	WTGCGNACCA	YGGCANCCTG	420
GGTATGGGTG	RAACASTSG	CGAACCGGT	CGCGTGGC	TGGGTGATY	GTGGGNTCCG	480
GAATTGCGG	GATTTACGG	TAMGTTTAA	CGAGGYCGG	TGGGTGCTY	CGAACCGCG	540
ATCMWCGCG	TACCTGTAA	AATGTTTTG	CGTGAACCG	AGYCAAAAA	YNTGNTGCGN	600
CGCMMGGCG	YTGGAAKT	CAATGTTG	MACGGCTNG	YTGGAATTT	TGTTGAGCG	660
YGGGCAAG	AGACCGGAK	YGTGAAYTG	WTAGCGTCT	TGCGGCTTA	ATGTYCTY	720
ATAGCGAA	CGGNTGWTG					

(A) LENGTH: 1166 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

x1	SEQUENCE DESCRIPTION	SEC ID NO:
		323

INFORMATION FOR SUBJ TO HQ 101

A LENGTH 1210 case 0010  
B TYPE 000000 0000  
C TRANSFER 000000  
D 000000 0000

## A. SEQUENCE DELETED FROM PAGE 3 OF 4.

NONSHNNHNT	ONTAKTOWN	TOTNCACCV	NGMTOWATG	GGGCGGCCAA	NGTTGEMNA	20
AGATATCTON	AAATGCGCA	ANATGTCTTT	TGNTAKTCT	GGGCGGNG	TAGGCGTAT	10
ATGCGGCTGG	GTTATATCAA	TGCGTCTCG	AGGCGGACG	GGGCGGCGA	TGAGGCGGCT	10
TTGCGGCGCG	GAGGTTATAT	TAACTGCGG	TGTTGTTTATA	GGGCGGCGA	GGGCGGCTTA	24

TCCCKTICGG	GGGCGCGCCN	AAAAACCACC	AATYCCGYTG	GGGGTGKYCC	CMCAGGCSGT	480
TGCTYCGNGY	CACCTGGCCA	AAVYCCCAWT	AKATTGGGTG	SCYCKTSCGG	TTSYTGGGGY	540
CAATTACCCC	CNCGGNAAA	GRRAAANA	ATCNTCCNTT	TGCTCGGYCA	YCITTTMTTG	600
SAAAAGGGGC	ATGGCSCGGT	TYTTTTACCT	CAAYCCCCNA	NCANTWACCT	YTCCSCCGG	660
GGGNCANAA	CGSTTNGCTC	CGSGGNAKCC	TKGTMCCCGN	ATCNAAAGGC	CNGAATTTGG	720
TYYSSTYCN	ATTWTWKKY	CCCCWNTTG	YAAAAKCCA	AAASAKCCCK	YCNCAMMYKT	780
NGGGGTYS	GCKNYCTTK	SNMTTAAACC	CYCCCCAAAA	YVNSGGGKKT	TCCGCYNSAT	840
KCCACCNCCK	GNGGGGGGNA	SAAAAAAAY	TTTYCCSAAA	ATCCCACCYY	TCYKTKSTRY	900
AMACCCCTT	TYMKKAYTC	CKYGCNATTC	SGMTTCWAAA	TYCCGYOGCT	TNTTCCCCCK	960
SCYTMAANTA	AAANAANGGG	NKTTTYCTY	CCCNAAAMNCM	AWTMGGGGKS	KCCATTCTGG	1020
AKMAAAKAGN	KKMTKNNSA	AANCCNCCCC	CTSTYTNYYT	NKTNMNCCKC	CYGGKGNKGM	1080
SWSWYNTTCT	NCCCRCCCC	YNYNKTGANA	AAMMNCYCCS	GGSTMCRNAN	ASNMTTTTCK	1140
STSTNMGCC	KMBASNANAN	MCAMWKWYCC				1200
						1230

# 2) INFORMATION FOR SEQ ID NO:325:

## 1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1022 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## (1). MOLECULE TYPE: Genomic DNA

## (2). SEQUENCE DESCRIPTION: SEQ ID NO:325:

NGNBSGKINA	TMAKWTCTC	ACSSGGTCTA	TGCGGCGCAW	CTMGTMAASA	GATCTCNAAY	60
TGCGCAMNAN	GCATMTMMC	CATATATAAC	CATTGCGTCS	GYWTSCAWCT	CRAAWCTGTC	120
TTTBSKGGGG	TTKTACRAAG	GTGGMWTGYT	CWTYCTTRAA	SCCCTCRATC	TCKTKTATYC	180
GTGGGGGTTC	ACTTTAACSG	RATKSTTSCC	TTKTAYCATT	RATGCAAWTA	WTGGYCRAWT	240
TTTSCAGGGC	RACGGGNYCT	TTTYCCCGRA	BRACAATNGA	TTGGAWYCGJ	TYCGCRAGGC	300
TGCGCACCAR	ACCGGGCNC	AAAGGYCCGC	GCAAWTSCCT	JGKTCAAAAA	TGGTGCRAAC	360
GAAMCNATGG	CGGGTTTAC	CGCAGYTAMC	ACAAKAAAAAT	TCCGVTGGCT	GCACCAWNNIT	420
TTTCTATCWY	GWYCCCGACG	TTTAAATTTGK	VTGCGGTATT	GCCTKCGTGC	CTCRACAGCM	480
TGCGGCKTCA	AACCTGCGGT	GACTCCAACT	GGTCTGGYCG	AASGGGGGYT	CAMCGGACAA	540
AACCGCRANW	TGCGCAAAAT	TTGNCGGCCC	TGCGGGAAN	GKTGATMTT	TGNAACCSA	600
MGSGGNNYTW	NAACCGTGAA	CGSSGSNKGA	MYNSCGSGGA	ANTTTTCCCT	TYNGGGCGRN	660
AAANCGTTTT	AAGGTACCCC	KGGNGGGGKG	GGGYTTGGG	AAACCAACCC	CKATTGGKTT	720
TGGAATNTT	TKNCCCCCA	TTGNSGGGGG	GGGCGGCAMT	TCMCGTTTTN	TCMSTNTTNY	780
CTVYCGCAAT	TNYTCCCGCG	SAAYYCGGGY	TYVYCGTAA	NCGCGMNGG	WYSTGSHAF	840
CTPATMAWNT	TNSTTTTNYC	CGCGGNNCGC	CCYAKGNT	NTNTGAAAMA	AAAKCGGCTC	900
CTNNMYMWY	CTGNGNRTT	TNCGGSSNMT	TYMAAAMMAN	CGGNNWTVY	CTGNGGCTNN	960
CTVYCGCAAT	TNYTCCCGCG	SAAYYCGGGY	TYVYCGTAA	NCGCGMNGG	WYSTGSHAF	1020
CTPATMAWNT	TNSTTTTNYC	CGCGGNNCGC	CCYAKGNT	NTNTGAAAMA	AAAKCGGCTC	1080
CTNNMYMWY	CTGNGNRTT	TNCGGSSNMT	TYMAAAMMAN	CGGNNWTVY	CTGNGGCTNN	1140
CTVYCGCAAT	TNYTCCCGCG	SAAYYCGGGY	TYVYCGTAA	NCGCGMNGG	WYSTGSHAF	1200
CTPATMAWNT	TNSTTTTNYC	CGCGGNNCGC	CCYAKGNT	NTNTGAAAMA	AAAKCGGCTC	1260

# 2) INFORMATION FOR SEQ ID NO:326:

## 1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1033 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:326

```

NNCGNNKNTA TAMAYCWYCT NCACCSGGGA TCWATTGCGG CCGCAATCTT STMAASAGAT    60
CTCKAAYTCG GCAMGANCCG CAWCTATTTG KGTGRASCGC ACCAGCGRGA CCTCGCSGKT    120
CKTTYCTTGC AGRGAGGCGC TGGGTGGCRC CGGIGGCAAT GCCAACCGCC CCCCAAAACN    180
CCGCCAAATMY CRAAAAACAA CCCSGGGSTA GKTCCSGGCC GCCAAATMAA TAACCGTKIT    240
AACKCAGGCN ACGGCCAACG GGYCCCGCCC AACCAAGCNA CTTCCCCSCC NATAGGYCCG    300
GTGGGGGCTG CCKTATYKCC AASTCGTCAY CTCNACGGGM CGGYCCMCWT TCCGCCTCAT    360
CGGTCTCTCC TTMATTTTC CRTCCACYKG GCGGGGAACY TTTTNYCNC CTTGSCMAN    420
CACCNAGGY CNAAAATTNC CMTCCCKYG SNNCAAAYGR GATTGGGGTY CGKXTTTTNT    480
TCNMCCMAAC CCCNTTTNA CGCCCCMATC CCYTWATACC CCCWWMCMNS ANGKTTGNSA    540
AAKTNNCCCC AAATRCAAA MTCTTCGCC NTTMTWMCY YYCCTTTCCC CMCCCNAAA    600
GGSCCRCCYY TCGGGAANTY TCCCNCAA AWTCAWCCM TTTCCCNCCA AGAAWTTCSG    660
SACTCCTTIN TTCNGGNGAM ATANATYTT YCKTNGGSK TCCGMTNC AMMAATNTCC    720
RGGGKAAMCC AGKNTNNTCC YYYCCCAA NNTYCCYKG RMCYNYCY TTAANRASE    780
SAACCCSGG GKCYNCNCS TARCCCCAM KAAATTTCC CCGSKITTC TYNNKKMRW    840
GCCCCSAAM ACTMTWAYTT TCCCKCGNN TTSYCKCS KCAMWMMTG KKNCTTTTT    900
YCSMATAMA CTNGGKCT NTNYGSGG CMAAANAAG CGCGSTCTN TTCWMAMACA    960
YNTSGNMMA SAAKAKWATA AWWNTRKYYK TKNCCCNCC CKKCTTSNN TNKCCMSKS    1020
GGGNNNNKKR CMTCCWNC CCCCCANA CCKWATMCC CCGCKCCGM NCMNTTTKT    1080
CCC

```

1083

## (2) INFORMATION FOR SEQ ID NO:327:

## (I) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1069 BASE PAIRS
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

## (II) MOLECULE TYPE: GENOMIC DNA

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:327

```

GGGGNNKYAT MCAYCWTCT YACSGGGMND TATTGCGGCC GCWYTNCTM CASAGATCTC    60
GAAYTCGGCA MGAATAAGW GATGTGCTG ACCTTMCCE GCGGGACGER ACCRACAAAG    120
RAASCAGGCC ANAATATTGG CCACAKTGG TCACATATT ACCCAATTMT AYCAGGGAYT    180
MCCATTCCGK GGACCRACC CACAATCCCR ATSKTGSTTT GCRAACCTR ACCGTCCCA    240
MYTYCGCCRA STGAACCA GGCRAAAAA CGGCCRAAWY CTCGCCCTGA NTCCCGCTCS    300
GCGCNAATAA CTAGGCCCAT TKAACGGAAC CGGNGGCCSC NANTTGGCCA ACAGGTCTTR    360
ACAAAGGGGG CCASYYCGG CCGGTCCCW TTYCACNCC TNKTCTCKTG CGGAATYCGG    420
WTCRATNYC CWTGGGCTT TKTXYCKYC KYCGGTNCA AWICTNCGTA TNCATRGKG    480
TCCCTAAAT SCANATDGG GEFYCCATT NTYGGNTH NATTTAMMAN SRRCGGTCT    540
TTCNTCCRA AACCGSNTG GCGNNMCA AAAAAATGAT AATAATSE YGCTITCAA    600
ACCGCGGCC CCATTBRWT CGGTCCANC CCGCNGSGT TARGTGGGA AFIYTNAM    660
YCNARGCCCT NATTTGONA AAAACCYCY GGYCTEAAA CNYTTTTT GSKSNTCCG    720
GCTGRTTSC CAAACCAA ATNTNYGG GGYCKTNA ACNCGGYCR RCISGAART    780
TTTTYGSTTC AACCCCAAC TTTCAASC NTTTYTYT TCCSGCSN TNSSSGSNT    840
RSSNNTTCY RAKKCNEN GNCGGYCYN CCCCMTTT CTTTTTTT CCGTNNAM    900
NGKTCTTCA AASMCCDCC SCGCCNSAA ACCCCCTNA GTTTTTYMA AANNWYNGN    960
NNCCCCCCC MGAATAAAY YCSCCGNPN ACSMSNGSA MCGCGGSGN NTTRKTTTT    1020
TNCMSGYCCC CORMASYTT TKAMAWRE GANNMTTY TNRGNNY

```

1069

## (2) INFORMATION FOR SEQ ID NO:328

## (i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 1210 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:328:

```

NGNGGGGKWK MATACATCWT TCTTCACGSG GGATCWATTG CGGGCCGCAW TCTNGTMCAA      50
SAGATCTCGA TYTCGGGCGAM NACCCACCWC TCCRAAAAAA ACCCRAAWCT CGGGSKCTYC      100
GARAAGTGTT GCGCGCKTTR AATTTAACAA ATTCACTGTC ANAGTGTCAJ GGCKTTACWT      150
YCCCGGCAAA GGGGCCACAA CCTGCAGRGA SCACYCRATG GKTGYTGKTS CNGCGGGGCGG      200
CGCGKTNAAG GGACCTGCTT GGGTKTGCSG TMCAAAANATC WYCCGGCGGGT YCGCTGGRAT      300
MCNCAGGGGT CTCAAAAAAG CGCAAACAGG CACSCCANCC NTTTTAGGGG GTTAAAAANGA      350
AAAAGCGCTG ATGCCCCCAA GGGGGGCGCGT NCTCAACCTT CCGTCTGTCA ACAACCCGCT      400
CTCTCKTGCC RAATCGGRWT CCRATNYCNC CWTGGCCTTK TCKYCTYCTT CGGTACCCAA      450
ATCTGCGTAT CCTATATGTT CCGCTAAWTT CCAATCTGG GCTGTGCATT TCGTTGGGCT      500
TCCAAATTTA CCANCAAGG TTTCTTNCAT NCCAAAAACC GNTKGSKCKC NRACCCRAAA      600
AAATGAATAA TAATAANNGG KNNNTTYCNA AGCNCGCCCG CCGNATTCCA TYSNGTTTCA      650
NMNCCGCCAG NGGKTAGGSK GGGAAANYYC TCMACCYICA ANCCCTWARS TTTTNGRAAT      700
KAAACGCTYC YCNGGGTJWW TYMAAAAAAMA NTTATTTGGN NGNTTTCGGG MWNCKRKJST      750
SCCAAAATCC MAAATANTTT YTTGGTYCNA TWAAAAAMCG YGNCCMNCCG GGAAAAATTT      800
TCTTGKTTSA ACCCCAAAAC YTTTTCHNAA NCSSKTTTTY CYTCCGCGCG AMNWTGGGYS      900
GGGNAFKGYG SCYTNCTTCA TKTKYTYMTW CMGGGGGGNN MKMTGMMCCG CCMTTTTYYCY      950
NYWRTTTTTN KCCCCNTNMP NNRAANNNGN YTCNNANAA AAGCNCGCCG SCCKNCCENA      1000
AAAAWCCCN NNNABAKTNT TTMKANNNMN SKCNKNGKY YCCCCCGWC YNMNNAAAAA      1050
AAATMYCNCC PASANMCASM NMGGRGNPSG CCGCGCGGTT NNNNTMTTNT TTTTTCGSA      1100
GACCKCCSCG MNNANMKNCY CTTTTYCHT NNGNNGNGNN GNGNMNCKCC CENAGAAMWF      1150
CTCTCTCKKS
  
```

## 2. INFORMATION FOR SEQ ID NO:329:

## a. SEQUENCE CHARACTERISTICS:

- A: LENGTH: 1105 base pairs  
 B: TYPE: nucleic acid  
 C: STRANDEDNESS: single  
 D: TOPOLOGY: linear

## b. MOLECULE TYPE: Genomic DNA

## c. SEQUENCE DESCRIPTION: SEQ ID NO:329:

```

NCCGSHNNNA TMCATCWATT GACCGGGMT CWATTGCTT CCGAACTNCT MAAGAGATCT      50
TGAATTCGGG AAKANACAGC AGCGCGCTGT MTATACACCG CAAATCTTCT GKTGCGCAAA      100
ACCGAGACCG CCGCGCGCGG GGGYTCCAAC GCKTTACYTE ACCCGCCAGY TCACTTTTCA      150
AAJCGGTGYT RAGCGCGCGA CTCACGWTAA ACCGTTTAKC CAAGFAWYTC SKTGGCGCGG      200
AATGACCTGY TCTGWTGCTC CTGWCCTGTC CTAGCGCGCG CTANCGCGG GTGCGCGGYTC      250
AMGACNCGCT CTCTATCTC CTGWTGCTC CTAGCGCGCG CTANCGCGG GTGCGCGGYTC      300
AAATGACCTG TCTGWTGCTC CTGWCCTGTC CTAGCGCGCG CTANCGCGG GTGCGCGGYTC      350
  
```

ATTTCSGRAA	SAACCCCTNY	CCCGGGTTTT	YCGTGGCTMG	GCSCAANACC	CGCGGGAATC	660
AAAAASGGTC	GGNCAAANGG	GCMAAACCCT	SACCCMACTT	WTTCCRCTTN	GGGGGGSCWN	720
CCKNGTTTTAA	AWKSCCTCYV	CTSCCCAAAY	TGGGKCMAAA	NNGRKTTGGK	TTNGGCNACC	780
NTTTCGGGKC	CCGGGKGKKG	WGKYCTMNMA	CGTTTTNTTTT	SCCCCYKAAA	NYSCCCCCCC	840
CGSSCCCCCG	CCCGGGGGGA	NNTTTTTAMA	GKTTYCCCCC	CCCCAMAAAA	ANACCCCNVC	900
CCGGGSCCCT	TTKRWAAAMN	KCTSCCCCN	GNNGGGGKCM	GGKTTATMT	NNNCCSCCCC	960
TCCGCGSAAA	AAATAKMTT	SYCCCCCNC	CTCCCKNCKNR	GKAMSMSCGC	TCCCYCTCNC	1020
GCNKNTWAAN	ARSNCKKNN	CNCYKCCGS	NSNGKCNWCD	NCCSTSSNCT	NKGCKCKNCN	1080
KAAANAAYNC	NGSMSTSSME	CNKCC				1100

(2) INFORMATION FOR SEQ ID NO:330:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 936 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

11 MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:330;

NGNSNSKNNN	TAMAYCWYYC	TSCACSNNGA	ACWANTGCGG	CCRMAWCTNS	TMKASAGATC	60
TMGAAYTCGG	CAAGAGCGGC	AAGAGTGTGT	GCATCTGGTC	ANAGTSTMMA	CRGGGTGCCC	120
CGGCTGKGTR	GACCAACMCAT	NTGGGRACAC	CAAAACCCCTC	CGGGGYCACC	CGCKTCCCT	180
GCAAAWYCCT	CCAGGCCACC	TCRAACAAYW	CTCTCTGCAA	CGCARGCCGT	TYCGGCGCCC	240
RATCTTGCKT	CASYTCGCK	TGCGGTGCCC	AAGKTACTGG	CCAYCAAAA	CGGCTCCGG	300
AACTCRAACKT	AAWTTGCGC	AAATTCNTTC	CGCTGCGCCT	TGATAAATTT	NTNAAGCCAC	360
CGCAAMCCTV	CGGGCKTCTC	CTCTCTGCGA	ATYCGRWTCG	RATAYCGCCA	TGGCCNTNKT	420
KYCTYCKYCS	GTACCCAAAT	CTTGGGTATC	CTATANTYTC	CGAAWANCRA	AWTCTCGGCK	480
KTCBATKTC	TGSGKTCCRA	ATTTAMMACA	NCGCTTTTCT	TCNTACCCAA	AACGCTTCGG	540
CGCCACCCRA	AAAAGGATAA	TAATAAKGTG	CGWWCAAAA	CGCGCCCCCG	ERTTCAYTCG	600
GTCCARCACC	CCANGNGGT	AGGTNGGAAT	CTTMAACCTC	CAACCGATAA	CTTTNSGNAA	660
AAACCCCGCN	GGGYMYCAA	AMMCTTTTTT	TGGMTTCGGG	TCATKGYKCN	CAAAACCAAA	720
CTTTCTCGCT	CGWAAACAA	CGCCCNCGCT	MAAATTTTTT	GKCAACCTCA	AACCTTTKAM	780
CTNNNTCTCY	CGCCNSACAA	TNGSGGCKN	NGGCTNTCTT	CTTTTTYNNA	CGGGGGRRC	840
GNCCCGNAAN	TYCCNAANKG	NKCCCGSNMA	AAAGAGANTT	TCMKAAAAA	CGCCGCTGCC	900
MAAAVACCCC	MAAAKWTTCT	AAASMSCTNG	CTCCCGC			930

## 2 INFORMATION FOR REQ " " " " "

## SEQUENCE NUMBER 10112

- A LENGTH 140 base pairs  
B TYPE nucleic acid  
C STRANDEDNESS single  
D TOPOLOGY linear

## 4. MOLECULE TYPE GENOME DATA

VI. SEVEN IS DESCRIPTION: 30; 17; 11; 11

[illegible]

SCGGRAASCG	GTGCCAACCC	RAAACNCKTT	GGGCACYCGG	KTSRACTTTA	AASGGTAATC	300
TCKTCCTCCT	GGGCTATGGT	GGGCCACAAA	CCTSYTGCCG	WGGGTCTGGC	CCTGGGYCAC	360
CGYCRCNTTT	TATNTNTCK	YCTACACNCT	TKGGTYCAAC	CAACCCACTT	CACMAAATTG	420
TTTTGGGKTG	GGGSSGCCGG	YTGTNNCCGK	TAATAATCSG	NTGKTCSGCC	MYCACC GGWA	480
CCATANCCTG	GGCGGCSCTG	GCAAATTTCC	SAAATCATYT	CCTTCTGRAC	CCCCACAMRC	540
CTNSAAATCC	GRATCAATNC	CCCNKGGCTT	NTCYCTCTCN	GTRCCCAATY	TGGTTTTCTAT	600
RKTNCCCYAA	TSCAATTGGS	TTYCCRTTSC	YGSTTCCAAN	TTNACAAMAS	GGTTTTYTCMT	660
ACCAAAACCC	NTGGSCCNNA	CMNAAAAKNA	RAAAANAKGG	KCTTTYAAAC	CCCCCCTAT	720
TCAWYCGGTN	CMRNWCCCCG	NGKAAGGKGN	GAAAYTTHRA	CCCAANCCMT	ARSTTSGNAK	780
AAACCCCYCG	GGGTSMCAAA	MKNTWTTSSC	CTTCGGMCTT	YCCAAATMSA	AAATYYTCKK	840
KRMNAAAAMC	YGNCCCCSAA	ANATTTTTGT	NAAMCCCKMA	YYTRTTWMCC	WTTTTCCYCC	900
CCMCNNSNSG	GNTNCCCTTY	TYATTTCYMM	MCRNNSGACN	CCCCMNTYTT	TWTTCKCWCN	960
MMARGSNNTT	RGRMMNMNCC	CCNCCCCNAK	MTCCNCAAAK	NTTTNAACN	NNKYCKCCCC	1020
CCCMWMNKNC	CCCCMNCMTT	TM				1042

## (2) INFORMATION FOR SEQ ID NO:332:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (E) MOLECULE TYPE: Genomic DNA

## (F) SEQUENCE DESCRIPTION: SEQ ID NO:332:

NNSGSGMKKK	ATAMATCWTT	CYSYACCSNG	GMTCWATTGC	GGCCGMAWTC	TNGTMAASAG	60
ATCTCGAAYT	CGGCAAAANAK	ACGCMAYGTC	AAGTGTRAVY	CGGTACACATA	TCMTCCGCGNG	120
TCAACMCCAA	AGCCGNGTCA	CCGYCTCCTT	GGGGCCCCAC	CCCCATCGGT	RATGCAACYT	180
CCCGCGCCAC	CCYCAAAAGG	KTCWTTTRAG	CCCTAAAGGT	CAMCAATTCC	TRAGGTVMCN	240
CACCGTTNNT	TGGCCCCCCC	RAWTYCTPAC	CCGUNAATWC	CGTAATCGGR	AATTTGGGCW	300
CCCGCTTGGG	CAATAACKTN	TTGGGCAAGC	CCGURWTCYC	NCTGGCCGGA	ATTCCGNCAT	360
CCCTTTAAGC	CKTGRACCGT	TTTCCCGGHT	CCCTAAAYTC	CTYNTTGGGC	CCCTTGGGCC	420
CPNAGCASYV	CCCTAACCGV	CMCCAGGCAA	TACCKTTGGC	TTTRAACCCAC	CGGRATNAAY	480
TGKTACCCAC	YTCMASSOTS	CTGRANTTRK	TNTGNTGRAA	AANMCCACCN	AACCCGGNTT	540
RATCTGCTTC	MTCAACWTTT	CCCGGCTTGT	CCGTTTTTGR	AAYCTTNATC	CMTYCAAAAG	600
CTCTAMTTTC	CCAAANPAAAT	CCGYTTGCCA	CCCTGGCCCG	CGCTGGTTTT	CGMWCCTTRR	660
AMATCCNCCG	CCGGGSAAN	AMTTSGGNTT	CCGCGGCTTC	CCGCAATAT	KCTTGGNCCT	720
MAAATTTGSS	GGGATCCCCN	CCGNAYCCCG	CCCTKCGJSH	CCGCCAGTTT	CAACAATTTC	780
CCCGCTTCCA	AAACGCGGNC	CCGGGGGTGG	CCCTGNTTTT	CCCTMYNNAAA	AAGKGTTCGN	840
CCCTTTTCCG	CCPAANTTCA	CCGCKNKTNT	CCGCAAACT	CCYCAANTTC	CAACCTTTTA	900
CAAAAANTTY	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	AAAAAAAAGNK	960
CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	1020
CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	CCCTTCCCG	1073

## (2) INFORMATION FOR SEQ ID NO:333:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 141 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:333:

GNSNGNKNTN	TMCAVCWYCT	SCACSGGGTC	TATTGCGGGC	GCAATYINGT	CKASAGATCT	60
CGATYTCGGC	AMNANAARTG	TCGTCTGCAA	TTTCAGKKTG	GTCKTCAAAY	GGGCCAGGCC	120
GNGACCRACA	CCCTGNGTCA	CCCCAAAAC	CAACAGCWTG	AAATWTCAAG	GCCRAGGCSC	180
TRTCAATYCC	CRASCAKTTA	ACCGTKTCCW	TCRAAGGTGC	CRAACCAGGC	ACCCAGYTCA	240
CCGCCSGGCA	AWTCGCGCTG	CCGGCCGGTN	TCAGCCTGAT	TYCTGACCCT	RWTCTGTSGG	300
TGGYCAMCNT	GGTGAAGGCC	CWWCCGCNA	AGAACTGGAG	GGCRAATTCC	CAGGANCCNA	360
GRAACCCNAG	GAACCCGCGG	TAKAANCCGG	CRAAACCRAG	GCCGYTGGCN	ATTCCNATTA	420
NAMSGGTTTG	CRACNTGGCC	RAACCGTTTT	CTTGGTCCGC	CTCGGCAACC	CTGGACCANT	480
TACCCCKTNC	CCGGNMCMAC	CYCGGGTNCT	TGKYCCCAAT	NTGCTCCCGC	GNRANTNGGC	540
CNAATTCCAG	GGCNCCANCT	TTCCGGCCCN	AATTCCCYTG	GTTAATCACC	GGGCNCCCT	600
GGTTTTGGGC	AACCCCNCTG	CTTMTTAAAA	CATTCCGSCC	CAATGGGNC	STTGGSAAT	660
TCCTNYCGGT	GGGGCGGGCR	ANMYTTCTCT	YCCCNAAASN	CTTAMYCCAN	TTCCGSSNTCC	720
CGGKCAAAWS	NGGGGGGGNA	AAGGGCCCCC	CGGNTSCKCC	GGGGKKGCC	CYGGKTTCAA	780
AANTTTCSGG	CKTSTMSCCS	NUTCSCCCCC	CGGCCAAGRA	CCGNGGTTTT	TTTTGAACC	840
KCMANTCOSA	AMCCGCCSSC	CCMAAACCS	GCCTHAAGWR	RAYTTNKSCC	CTNAAACSGG	900
CCCCCAKTYT	SGGKTTCCNC	CNCCSGKKGT	CCMTSTTTMM	MRCCTTTTGN	GNKTTTTTAN	960
MGSCCTTNNC	CACCCCCYCK	GGGKCSMNNA	GAATMYWKC	CNGGGGNAN	RSCTCCCTNN	1020
GGGKGGGGKG	MGAGYSCCKT	CTKCGGNCNN	YKNTTTCCCC	C		1081

## (X2) INFORMATION FOR SEQ ID NO:334:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 986 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (E) MOLECULE TYPE: Genomic DNA

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:334:

GNSNGRINKWN	ATMCAYCWYV	NTSCACCCGG	GMTCHATTGC	GGCCGCAWKY	WNGTMAASAG	60
ATCTMGAAYT	CGGCACANAG	CGGCACAGAG	TGTGTGCATC	TGTGTGANAG	WTGTCAACCC	120
GTGCGCGGCG	GTGGTRASCA	CMCATTGCGR	AACACCAAA	CCGTCCGCGG	GYCACCCGGK	180
CCGCCTGCAA	AAYCCTCCAG	CCGACCCYCA	AACAAYWYCT	CGTGCAACSC	ARSCCGTTYC	240
CCGGCCGRAT	CCTGGKYCAS	TTGCGCKTGC	CGTGCCGCCA	GGTACTGGCS	GWYCRANACC	300
CTYCGGGRA	ACCHAACTA	AATCTTTCGN	AATTTGCHTT	CGCGCTGCGC	CTTATMAATT	360
CTTTAAACCA	CGCAACCTV	CGCGCTGCTG	CTCTGCGRA	WTGCGPWTTC	CACTYCGGCA	420
CTGCTGKTE	KYCTYCHYCC	CTMCGCAAT	CTTGTATATC	TATATTGTTC	CTAAATGCCA	480
CTGCTGCTG	TGCTATNTCT	CTTCTTAAA	CTWAMANGAG	NGGTTTTCTT	CTTCGNAAAC	540
CTGCTGCTG	CAACCCNAAA	AATCTTATA	ATAATGGTGC	TCTCAAACCC	CGGCGCCATY	600
CTATGCGKCC	AMCCGCCCGN	CTCTATPKC	CTAATCTTMM	AACCCCAAGT	CTATAASNTTG	660
CTAATAAAT	CTGCGCCYCA	CTAATAACANY	CTTNTTGGNY	CTNTTGGGMM	CTATGGCTNN	720
CTAATAAAT	AATCTTNYG	CTGCTAATAA	AAMMMSCGYC	CTMCCCGAAA	CTTTTTTTGN	780
CTAATAAAT	AATCTTTTTT	CTAATCTTAN	CTTCTCTTCC	CTCTMANTGG	CTNSGGARTKT	840
CTAATAAAT	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	900
CTAATAAAT	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	960
CTAATAAAT	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	CTGCTTCCAA	1020

- (A) LENGTH: 1074 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

```

NGNGGGNKRN ATMMAYCWCT SATYYACCSN GGMNMWATTG CGGCCRMATW CTNGTMKASA      60
GATCTMGAAG YTCGGCAAAG AGYATKCTCG GGGGCCAGAT TTNTGGCCCG CAACCGCCCGC      120
ACTTTGCAYW TCAACAKTCC SGGTGCCCA AAAAATCWT ACCCCCATMC TYCKTGCASM      180
ASYTGCGCCC RATTRAACAC CCGGCCGGCW TGCTGCGCCA GGTATTYCAS CAGYTCAAAY      240
YCTTTKTAGK TAAATCCAG CGGGCGGCCA CACAGCCGGG CGGTKTAGGT GCCTYCRTCA      300
ATMACCAGCY CGCCAGGGY CACCTTGCCC AAAAYCTCCT GGGTCAGCCA AATTYCCGCS      360
CGGGCCAACM ACCANCCSCA TYCTGGCCTC AATCYCACCG CGCCCGGTGY TAAAMMANMA      420
BRATCTCKTC MANCCCCCAN TCAGCSYTN AAGCMACAGC CGCCCTTCTT CAMACCGCCA      480
GTACCCGGWT CAACCGCCCG GTCAACCTCA ACAGGGCGGNC AGGCCTCCCC CGGANSAAAG      540
GTCTTACSCC NNYYAANAAAA MAAGNTCTGT TTTCCCGCTC CAAASNAAG AANCCCGSGC      600
CGGGCCTTCN NMMGGGTTTC GGGMANANAA AARCNCCGGN GGAACGNATC CGAAAMCTCC      660
CAAGTCNCMT TWAWAACYCN NNAACCCCC ANTTTTGGGA AAGGNTCCCC NTTMYCCCCC      720
TTTTASGKTS GGGMMYYCTY TAAAAAATT CCCCCAAAAG CCGCGGGAAG GGTGMAMCTG      780
GGNAAATTC CAAMCCNWKH TTTTYNGGT TMCGGGGGRA AATTYCNCCT CCYYNNNGGG      840
CSSGSNNNAT TAYGGSNMT TTTNNAWTM NSGKKTSAAM YNNKCCMNNN SNNMSMANNK      900
TNAMCKCCCN CCTCNGNGKY CCCYNQCCG GNAGNGGRAS MKCCNANMAA AYASGNTTNK      960
CGGAAMMCNN AATKGNNSC CGGASMCNN NNMAAATMT CNCNKCNSNN AANRGMRACN      1020
CCNSNSGMN RRGAAARMTNY CCCCCCGSKM GKGNKAAAAW GKYCCCCCCM AAAG      1074
  
```

1. INFORMATION FOR SEQ ID NO:336

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 1198 base pairs  
 B. TYPE: nucleic acid  
 C. STRANDEDNESS: single  
 D. TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

11. SEQUENCE DESCRIPTION: SEQ ID NO:336

```

NNNNNNNNNT MTAAATGTTT TTAATGTTT NNTTAAATTT GGTTCATATW TTTTGCASAG      60
TTTTCGAAYT TGGGAMGAGG AATTCGJJA LNNNCCGAGN NASTTGTGCT TGTGTACCCC      120
TTTCHGCGCC TCAAGCGCTCC AATGAGCAG CAGACTGCGN TCGCTTYTCT TTTTGGCGGC      180
TGGCTCACCG EKCTENWTTT PAAGGCWTCG TGAACCGGNT TCGGTTTTCT TAAACCTGGG      240
AAAWTGGCCA TCGCTGTGCT TCAATGGGNT TACGCAACCG CNGCCGCCAA TCTTTCTTA      300
AATCGGGYCC NTGCTGAGCT TTTTGAAYCT CCGGUSAAGA ACTGCTTCCC CNGGAYCTGC      360
TGGAACTTRK TGNAAATGCT GCANAKTCTT TONTAMGYCT CNGCCGAAGC NGAACCTACT      420
TTTNGGWANG TCGGCKKCCG CGGCTTATCA TCGCTGATCA ACCGGGAAGT GGYGNNSTTG      480
KGGGAAAAAG TCGCTCAAT TTYGGTCTCT TCTTGGKANT CCGGCGCTGY TYGCGNAATG      540
TAAAGCCMAE TTTTAANDCT TTYGNYCTT TCGCTTSTA GKWTTYCTG MGGANKAMNN      600
NNNMAWTTT TGTNCGGCT ATCTGCTGCA TTTTAKAJA ANACTYCCYX AGCTNTNTYS      660
TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT      720
TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT      780
TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT      840
TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT      900
TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT      960
TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT      1020
TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT TAAATCTCTT      1074
  
```

TATKSAAGMGG	TKCCGAGMK	CSCGTTTTKT	TKTGANAAMN	MSMRKNGTGT	CGMYTCTSC	960
GGGNTTTGTA	GAGTAKTGG	CSCSSMWGAC	WCSGCMCMNG	AGKNKTNNTS	VANTGARCGY	1020
MNNSKTKMKMT	MSCSCCGGNA	CGAGNGCCCC	CSANGMSTGY	NKGGMNNSNG	ARAKGATGGS	1080
GGCCNCGMNN	MGGGAGNMGA	SANNGMGMR	GGGGKTKGK	TKCSCCGGS	CSANGRAGAA	1140
SKTCNGSCGC	CGMGGKYGKT	KTKTKNKTGG	YSTCMMMM	NAGAAAAGAS	AGGCG	1199

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3572 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

11) MOLECULE TYPE: Genomic DNA

X27 SEQUENCE DESCRIPTION: SEC ID NO: 337.

CCATCTGATC	TTTGGCAACC	AGCATCGCAG	TGGGAACGAT	GCCTCATTC	AGCATTTCGA	60
TGGTTT3TTG	AAAACCCGAC	ATGGCACTCC	AGTCGCTTC	CCGTTCCGGT	ATCGGCTGAA	120
TTTGATT3CG	AGTGAGATAT	TTATGCCAGC	CAGCCAGACG	CAGACGCGCC	GAGACAGAAC	180
TTAAAT33CC	CGGTAACACC	CGGATTTGGT	GGTGACCCAA	TCCGACCAGA	TGCTCCACGC	240
CCAGTC3CGT	ACCGTCTTCA	TGGGAGAAAA	TAATACT3TT	GATGGGTGTC	TGGTCAGAGA	300
CATCAAGAAA	TAACGCTCGA	ACATTAGTGC	AGGCACTTC	CACAGCAATG	GCATCCTGGT	360
ATCCACGCGG	ATAGTTAATG	ATCAGCCCAAC	TSACGCGTTG	CGCGAGAAAG	TTGTGCACCG	420
CGGCTTTTACA	GGCTTCGACG	CGCTTCGTTT	ATCAGATCGA	CACCACCAGA	CTGGCACCCA	480
TTTGATT3GC	CGGAGATTTA	ATCGCGCGCA	CAATTT3CGA	CGCGCGGTGC	AGGCGCAGAC	540
TGGAGGTGGC	AACGCGAAATC	AGCAACGAT	GTTCGCGCGC	CAGTTGTGTG	GCCAC3CGGT	600
TGGGAAT3TA	ATTGAGCTTC	GCGATCGCCT	CTTCGACTTT	TTCCGCGGTT	TTGGCAGAAA	660
CTGTGCTGGC	CTGTTCGACG	AJCGCGGAAA	CGGTCTGATA	AGAGACACCG	GCATACTCTG	720
CGACATCGTA	TAACGTTACT	GGTTTCACAT	TCACAC3CT	GAATT3ACTC	TCTTCGCGGC	780
CTATCATGTC	CATACCGCGA	AAGGTTTT3C	CGCATTCGAT	GGTGTCCGCG	ATCTG3ACGC	840
CTGCGCTTAT	TGAGCTCTTC	TAATAGGAAG	CGCGCGATA	GTAGCTTGAG	CGCTTTGAGC	900
AGCGCGCGCG	GAAGGAAT3G	TGCTG3AAG	AGATAT3CG	CGACCAATGC	CGCG3CGACG	960
CGCGCTGCGA	CGATACCGA	CGGAAACAA	CGCTGATGCG	CGCGCAAGTC	CGGAGCGCGA	1020
CTTTTCGCGAT	CGGTGAT3TC	CGCGATATAG	CGCGCAG3AA	CGCGAGCTGT	CGCG3CGGTG	1080
ATCGCGCGCGA	CGATCGCTTC	CGGUTAGAGG	ATCGAGATCT	CGATCGCGCG	GAATTAATAG	1140
TAGTCACTAT	AGGGGAATTC	TGAGCGGATA	ACAATTCGCG	TGTAGAAATA	ATTTTGTTTA	1200
ACTTTAAGAA	CGAGATATAC	ATATGGGCGA	TGCTGATCAT	CATCAGCTGA	TGGA3ATCAT	1260
TGGGACGAGC	CGACATCGCT	TGGAAGAG3C	CGGCTG3AG	CTGCT3AGAC	CGG3CGCGGA	1320
TAGCTCGAT	AGCATCGGCG	TGCTGCGAT	GAATTAAGAG	GACAT3CGCG	TGGA3AGCGT	1380
TGGCAAGAT	AGTAGCGTA	TGAATTCGGA	CGCTGCTTTT	GAATTAAGAG	CGCGCGACCG	1440
TAGCGGCTCG	AAACCAACCG	CGCTGCTTTC	TGAAG3CGG	CGCTGCTTTC	TGCTGCTCGC	1500
TATTACCGCG	CGCTGCTTTC	CGCTGCTTTC	TGAAG3CGG	CGCTGCTTTC	TGCTGCTCGC	1560
CGTGTTCGAC	CTGTGGGCTT	CGCTGCTTTC	TGAAG3CGG	CGCTGCTTTC	TGCTGCTCGC	1620
TGAG3CGACG	CGTGTGCTTC	CGCTGCTTTC	TGAAG3CGG	CGCTGCTTTC	TGCTGCTCGC	1680
CGCTGCGGAC	CGCTGCTTTC	CGCTGCTTTC	TGAAG3CGG	CGCTGCTTTC	TGCTGCTCGC	1740
CGCGCTAGCG	ATCTCGGCTC	AGCGGCTGAA	CTA3AGCTTC	CGCTGCTTTC	TGCTGCTCGC	1800
TAAGCTCGAA	CGAAAACTCC	TGGCGGCGAT	CTA3AGCTTC	CGCTGCTTTC	TGCTGCTCGC	1860
TAGGCGAGAT	CGT3CGCTTA	AGCGGCTGAA	CTA3AGCTTC	CGCTGCTTTC	TGCTGCTCGC	1920
TAGCGCTTCG	GA333GCTTC	CTA3AGCTTC	CTA3AGCTTC	CGCTGCTTTC	TGCTGCTCGC	1980
TAGCGAG3CG	CG333GCTTC	CTA3AGCTTC	CTA3AGCTTC	CGCTGCTTTC	TGCTGCTCGC	2040

```

CGCGCGGGCT GGGTTGSCAT CGAAAACCCG GCGGAACCAG GCGATTTCGA TGATCGACGG 2280
GCCCCCCCCG GACGGGTACC CGATCATCAA CTACGAGTAC GCCATCGTCA ACAACCGGCA 2340
AAAGGACGCC GCGACGCGCG AGACCTTGCA GGCATTTCTG CACTGGGCGA TCACCGACGG 2400
CAACAAGGCC TCGTTGCTCG ACCAGGTTCA TTTCCAGCCG CTGCGCGCCG CGGTGGTGAA 2460
GTTGTCTGAC GCGTTGATCG CGACGATTCG CAGCGCTGAG ATGAAGACCG ATGCCGCTAC 2520
GCTCGCGCAG GAGGCAGGTA ATTTCCAGCG GATCTCCGGC GACCTGAAAA CCCAGATCGA 2580
CCAGGTGGAG TCGACGGCAG GTTCGTTGCA GCGCCAGTGG CGCGGCGCGG CGGGGACGGC 2640
CGCCCAGGCC GCGGTGGTGC GCTTCCAAGA ACCAGCCAAT AAGCAGAAGC AGGAACTCGA 2700
CGAGATCTCG ACGAATATTC GTCAGGCGCG GTCCAATAC TCGAGGGCCG ACGAGGAGCA 2760
GCAGCAGGCG CTGTCTCTCG AAATGGGCTT TGGATTGAGC TTCGCGCTGC CTGCTGGCTG 2820
GGTGGAGTCT GACGCCGCCC ACTTCGACTA CGGTTGAGCA CTCCTCAGCA AAACCACCGG 2880
GGACCCGCCA TTTCCCGGAC AGCCGCGCGC GGTGGCCAAT GACACCCGTA TCGTGCTCGG 2940
CCGGCTAGAC CAAAAGCTTT ACGCCAGCGC CGAAGCCACC GACTCCAAGG CCGCGGCCCCG 3000
GTTGGGCTCG GACATGGGTG AGTTCTATAT GCGCTACCCG GGCACCCGGA TCAACCAGGA 3060
AACCCTCTCG CTYGACGCCA ACGGGGTGTC TGGAAGCGCG TCGTATTACG AAGTCAAGTT 3120
CAGCGATCCG AGTAAGCCGA ACGGCCAGAT CTGGACGGGC GTAATCGGCT CGCCCGCGGC 3180
GAACGCACCG GACGCGCGGC GCGCTCAGCG GTGGGCGGAA TCGATTCCGC GGACCGCCAA 3240
CAACCCGGTG GACAAGGGCG TCGCGGCGAA TCGATTCCGC GTTTGGTCCG 3300
CGCCCGCGCG GCGCGCGCGG GCGAAGTGGC TCCTACCCCG ACGACACCGA CACCGCAGCG 3360
GACCTTACCG GCGTGAAGAT TGTGCAGATA TCGATCACAC TGGCGGCGCG TCGAGCACCA 3420
CGACCGCCAC CACTCAGATC CGGCTGCTAA CAAAGCCCGA AAGGAAGCTG ACTTGGCTGC 3480
TCCGACCGCT GAGCAATAAC TAGCATAACC GCTTGGGCGC TGTAAACCGG TCTTCAGCGC 3540
TTTTTTGCTG AAAGGAGGAA CTATATCCGG AT 3572

```

# (2) INFORMATION FOR SEQ ID NO:338:

## 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## 2. MOLECULE TYPE: peptide

## 3. SEQUENCE DESCRIPTION: SEQ ID NO:338:

```

Val Glu Phe Glu Ser Gly Gly Asp Asn Ser Pro Ala Val Tyr Lys Lys
      5              10              15
Asp Gly Lys Arg
      20

```

# 4. INFORMATION FOR SEQ ID NO:339:

## 1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## 2. MOLECULE TYPE: peptide

## 3. SEQUENCE DESCRIPTION: SEQ ID NO:339:

Figure 1. The effect of the concentration of the *Agrobacterium* suspension on the transformation efficiency of *Agrobacterium* strains. The *Agrobacterium* strains were incubated in the presence of 100 mg/ml of gentamicin and 100 mg/ml of rifampicin. The concentration of the *Agrobacterium* suspension was 10<sup>6</sup> cells/ml. The transformation efficiency was determined by the number of transformants per 10<sup>6</sup> cells. The data are the mean  $\pm$  SD of three independent experiments.

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

CTAGTTAGTA CTCAGTCGCA GACCGTG

27

(2) INFORMATION FOR SEQ ID NO:344:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

GCAGTGACGA ATTCACTTCC ACTCC

28

(2) INFORMATION FOR SEQ ID NO:345:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2412 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:345:

ATATGSGGCC	ATCATCATCA	ATATCATCTG	ATCSACATCA	TCSSGACCCAG	CCCCACATCC	61
TCGGAAACAGG	CGGCGGCGGA	GGCGGTCCAG	CGGCGGCGGG	ATAGCCTCGA	TCACATCCGC	122
GTGGCTCGGG	TCATTGAGCA	CGACATGGCC	GTGGACAGCG	CGGGCAAGAT	CACCTACCGC	183
ATCAAGCTCG	AAGTGTCTTT	CAAGATGAGG	CGGCGGCAAC	CGAGGGGGTC	GAACACCACG	244
AGCGGTTCCG	CTGAAACCGG	CGCGGGCGCC	GGTACTGTGG	CGACTACCGC	CGCGTCTGCG	305
CGCGTCACTT	TGGCGGAGAC	CGGTAGGACG	CTGCTCTAGC	CGCTGTTCAA	CGTGTGGGGT	366
CGCGGCTTTC	ACGAGAGGTA	TGGGAACCTG	AGGATCAGCG	TCGAGGGGCA	CGGTTCCTGT	427
CGCGGGATCG	CGCAGGCGCT	TCGCGGAGG	TCGACATTCG	CGGCTTTCGA	CGGTATCTCT	488
CGGGAAGGTC	ATATGCGCGC	CGCAGAGGCG	TCGATGAACA	TCGCGGTAGT	ATGCTCGGCT	549
CGGGAAGGTC	ACTACACCTT	TCGCGGAGTG	ACCGAGCAGC	TCGAGCTGAA	TCGAAAGGTC	610
CGGCGGCGCA	TCGACGAGCG	CGCAGGAGAA	ACCTGGGAGC	ACCGCGAGAT	CGGTGCGGTC	671
ACCGCGCGCG	TCGAGCTTCC	CGGACCGGCG	GTAGTTCGCG	TCGACGCTTC	CGACGGGTCC	732
GTGACACCTT	TCGCTTTCAC	CGAGTACCTG	TCGAGGCAAG	ATCGCGAGGG	TCGCGGCAAG	793
TCGCGCGGCT	TCGCGACCAAG	CGTGGACTTC	CGGCGGCTGC	CGGCTGCGGT	CGGTGAGAAC	854
CGCAACCGCG	CGATGCTGAC	CGGTTGCGCG	GAGACACCGG	GCTGCGTGGC	CTATATCGGC	915
ATGAGCTTCC	TCGACGAGCG	CGTCAACCGG	CGACTCGCGG	AGGCGCAAGT	AGGCAATAGG	976
TCGCGCAATT	TCGCTTTCGG	CGACGCGGAA	AGCATTCAGG	CGGCGGCGCG	TCGCTTTCGG	1037
TCGAAACCGG	TCGCGCAACG	CGGCTTTCGG	ATGATTCAGG	CGGCGGCGCG	TCGCTTTCGG	1098
TCGATATGCA	ATGATTCAGT	CGGCTTTCGG	TCGCTTTCGG	TCGCTTTCGG	TCGCTTTCGG	1159

```

AATTTTCGAGC GGATCTCCGG CCACCTGAAA ACCCAGATCG ACCAGGTGGA GTCGACGGCA 1380
GGTTTCGTTGC AGGGCCAGTG GCGCGGCGCG GCGGGGACGG CCGCCCAGGC CGCGGTGGTG 1440
CGCTTCCAAG AAGCAGCCAA TAAGCAGAAG CAGGAAGTCG ACGAGATCTC GACGAATATT 1500
CGTCAGGCGCG GCGTCCAATA CTCGAGGGCC GACGAGGAGC AGCAGCAGGC GCTGTCTCTG 1560
CAAATGGGCT TTGTGCCAC AACGGCCGCC TCGCCCGCGT CGACCGCTGC AGCGCCACCC 1620
GCACCGGCGA CACCTGTTGC CCCCCACCA CCGGCGCGCG CCAACACGCC GAATGCCCCAG 1680
CCGGGCGATC CCAACGCAGC ACCTCCGCGG GCCGACCCGA ACGCACC GCCACCTGTC 1740
ATTGCCCCAA ACGCACCCCA ACCTGTCCGG ATCGACAACC CGGTGGGAGG ATTCAGCTTC 1800
GCGCTGCCTG CTGGCTGGGT GGAGTCTGAC GCCGCCACT TCGACTACGG TTCAGCACTC 1860
CTCAGCAAAA CCACCGGGGA CCGGCCATTT CCGGACAGC CGCCGCCCGT GGCCAATGAC 1920
ACCCGTATCG TGCTCGGCGG GCTAGACCAA AAGCTTTACG CCAGCGCCGA AGCCACCGAC 1980
TCCAAGGCCG CGGCCCGGTT GGGCTCGGAC ATGGGTGAGT TCTATATGCC CTACCCGGGC 2040
ACCCGGATCA ACCAGGAAC CGTCTCGCTC GACGCCAACG GGGTGTCTGG AAGCGCGTCG 2100
TATTACGAAG TCAAGTTCAG CGATCCGAGT AAGCCGAACG GCCAGATCTG GACGGGCGTA 2160
ATCGGCTCGC CCGCGGCGAA CGCACCGGAC GCGGGGCCCC CTCAGCGCTG GTTTGTGGTA 2220
TGGCTCGGGA CCGCCAACAA CCGGTGGAC AAGGGCGCGG CCAAGGCGCT GGCGGAATCG 2280
ATCGGCGCTT TGGTCCGCCC GCGCGCGGCG CCGGCACCGG CTCCTGAGA GCGCGCTCCG 2340
CGCGCGCGCG CGCGCGGGGA AGTGGCTCCT ACCCGACGA CACCGACAC CCAGCGGACC 2400
TTACCGGCGT GA

```

## (2) INFORMATION FOR SEQ ID NO:346:

## (A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 902 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:346:

```

Met Gly His His His His His Val Ile Asp Ile Ile Gly Thr Ser
1      5      10      15
Pro Thr Ser Trp His Gln Ala Ala Ala His Ala Val Gln Arg Ala Arg
20     25     30
Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
35     40     45
Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu His Val
50     55     60
Ser Phe Lys Met Arg Phe Ala Gln His Arg Gly Ser Lys His Ser Ser
65     70     75
Gly Ser Phe Gly Thr Gly Ala His Ala Thr Thr Val Ala Thr Thr Pro
80     85     90
Ala Thr Ser Phe Val Thr Leu Ala Gln Thr Gly Ser Thr Leu Leu Tyr
95     100    105
Pro Leu Phe Asn Leu Thr His Pro Ala Phe His Glu Arg Tyr Pro Asn
110    115    120
Gly Thr Ile Thr Ala Gln His Thr Gly Ser Gly Ala Gly Ile Ala Gln
125    130    135
Ala Ala Ala Gly Thr Thr Asp Thr His Ala Ser Asp His Thr Thr Ser
140    145    150

```

180 185 190  
 Leu Lys Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile  
 195 200 205  
 Lys Thr Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn  
 210 215 220  
 Leu Pro Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly  
 225 230 235 240  
 Asp Thr Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly  
 245 250 255  
 Trp Gly Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val  
 260 265 270  
 Pro Gly Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys  
 275 280 285  
 Ala Glu Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp  
 290 295 300  
 Gln Ala Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser  
 305 310 315 320  
 Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala  
 325 330 335  
 Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp  
 340 345 350  
 Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile  
 355 360 365  
 Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala  
 370 375 380  
 Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp  
 385 390 395 400  
 Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp  
 405 410 415  
 Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala  
 420 425 430  
 Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile Ser Gly Asn Leu  
 435 440 445  
 Lys Thr Gln Ile Asn Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly  
 450 455 460  
 Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg  
 465 470 475 480  
 Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser  
 485 490 495  
 Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Ala Ala Asp Gln Gln  
 500 505 510  
 Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Thr Ala Pro Thr Thr Ala  
 515 520 525  
 Leu Thr Pro Ser Ser Thr Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro  
 530 535 540  
 Ala Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro  
 545 550 555 560  
 Asp Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro  
 565 570 575  
 Thr Thr Val Ile Ala Thr Asn Ala Pro Gln Pro Val Asn Ile Asn Asn  
 580 585 590  
 Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr

Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr  
 625 630 635 640  
 Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu  
 645 650 655  
 Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu  
 660 665 670  
 Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser  
 675 680 685  
 Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys  
 690 695 700  
 Phe Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Trp Thr Gly Val Ile  
 705 710 715 720  
 Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp  
 725 730 735  
 Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala  
 740 745 750  
 Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro  
 755 760 765  
 Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala  
 770 775 780  
 Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu  
 785 790 795 800  
 Pro Ala

# 2) INFORMATION FOR SEQ ID NO 347:

## (1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 34 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

## 1) MOLECULE TYPE: Other

## X1 SEQUENCE DESCRIPTION: SEQ ID NO 347:

GGATCCAAAC CACCGAGCGG TTCGCTGAA ACCG

34

# 2) INFORMATION FOR SEQ ID NO 348:

## 1) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 37 base pairs
- B) TYPE: nucleic acid
- C) STRANDEDNESS: single
- D) TOPOLOGY: linear

## 1) MOLECULE TYPE: Other

## X1 SEQUENCE DESCRIPTION: SEQ ID NO 348:

GGATCCAAAC CACCGAGCGG TTCGCTGAA ACCG

## (1) SEQUENCE CHARACTERISTICS

(A) LENGTH: 1962 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:349:

```

CATATGGGGCC ATCATGATCA TCATCAGGGA TCCAAACCAC CGAGCGGTTT CCCTGAAACG      60
GGGCGCGGGCG CCGGTACTGT CGCGACTACC CCCGCGTCGT CGCCGGGTGAC GTTGGCGGGAG      120
ACCGGTAGCA CGCTGCTCTA CCCGCTGTTT AACCTGTGGG GTCCGGGCTT TCACGAGAGG      180
TATCCGAACG TCACGATCAC CGCTCAGGGC ACCCGTTCTG GTGCCCGGAT CGCCAGGGCC      240
GGCGCGGGGA CGGTCAACAT TGGGGCGCTC GACGCCTATC TGTCCGAAGG TGATATGGCC      300
GGGCACAAGG GGTGATGAA CATCGCGCTA GGCATCTCCG CTCAGCAGGT CAACTACAAC      360
CTGCGCGGAG TGAGCGAGCA CTTCAAGCTG AACGGAAGG TCCTGGCGGC CATGTACCAG      420
GGCACCATCA AAAGCTGGGA TTAGCTGAGG ATCGCTGCGC TCAACCGCGG CGTGAAGCTG      480
CGCGCGCAGG CGGTAGTTCC CTTGCAAGCG TCCGACGGGT CCGGTGACAC CTTCTTCTTC      540
ACCGAGTAGG TGTCAAGCA AGATCGGAGG GGTGGGGCA AGTGGCGCGG CTTGCGCACC      600
ACCGTCGACT TCCGGCGGGT GCGGGGTJCG CTGGGTGAGA ACCGCAACGG CGGCATGCTC      660
ACCGGTTGCG CGGAGAGAGG GGGGTGGTGG GGTATATGCG GCATCAGCTT CCGGACCGAG      720
CGCAGTCAAC GGGGACTCGG CGAGGCGCAA CTAGGCAATA GCTCTGGCAA TTTCTTGTTC      780
CGCGCGGATT CGATCATCGA CGGGCGCGCG CGGCAACGG CGGAGAGCTT GCAGGCATTT      840
TACCGCATCG TCAACAACCG GCAAAAGGAG CGCGCACCG CGGAGAGCTT GCAGGCATTT      900
CTGCACTGCG CGATCAGCGA CGGCAACAG GCTTGTTC TCGACGAGGT TCATTTCCAG      960
CGCGTGGTGG CGGCGCTGCT GAAGTGTGCT GACGCGTTGA TCGCGACGAT TTCCTCGGGA      1020
GCTGCGCAJG GGGGAGGCTC AGGTGGAGGT TGTGGCGGGA GGTGCGCCAC AACCGCGCGG      1080
TGGCGCGCGT CGACCGCTGC AGCGCGACCG CGACCGGCGA GACCTGTTGC CGCGCGACCA      1140
CGCGCGCGTG CGACACCGCG GAATGCGGAG TGGGCGGATC CGACCGCAGC AGCTCGCGCG      1200
CGCGACCGGA AGCGACCGCG CGCACGCTGC ATGCGCGGAA AGCGACCGCA AGCTGTCGCG      1260
ATCGACCAAC CGGTGCGAGG ATTCAGCTTC CGGCTGCTG CTGGGTGGGT TGAGTCTGAG      1320
CGCGCGCAAT TCGACTACCG TTCAGCACTC TCGAGCAAAA TCGCGCGGGA CGCGCGATTT      1380
TGGCGGAGAT CGCGCGCGGT GCGCAATGAG ACCGCTATCG TCGTGGCGCG GCTAGACCAA      1440
AAGCTTTACG CGAGCGCGCG AGCCACCGAG TCGAAGGCTG CGCGCGGTT GGGCTCGGAG      1500
ATCGGTGAGT TGTATATGCG CTACCGCGCG AGCGGATCA AGCAGGAAAC GGTCTCGCTC      1560
ACCGCAACCG CGGTGCTCTG AAGCGCTGCG CATTACGAAG TCAAGTTGAG CGATCGGAGT      1620
AAGCGGAACG CGCAGATCTC GACCGGCGTA ATCGGCTGCG CGCGCGCGAA CGCACCGGAG      1680
CGCGCGCGCG CTGAGCGCTC TTTTGTGTA TGGCTGCGGA CGCTCAAGAA CGCGGTGGA      1740
AAGCGCGCGG CGAGCGCGGT TTTGGAAGCG ATCGGCTGCT TGTGCGCGCG CGCGGTGGA      1800
CGCGCGCGCG CTGCTGCGGA TTTGCTGCTG TTTGCGCGCG TTTGCGCGCG TTTGCGCGCG      1860
CGCGCGCGCG TTTGCTGCGGA TTTGCTGCTG TTTGCGCGCG TTTGCGCGCG TTTGCGCGCG      1920
CGCGCGCGCG TTTGCTGCGGA TTTGCTGCTG TTTGCGCGCG TTTGCGCGCG TTTGCGCGCG      1962

```

## (2) INFORMATION FOR SEQ ID NO:349:

## (1) SEQUENCE CHARACTERISTICS

(A) LENGTH: 612 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

Met Gly His His His His His His Gly Ser Lys Pro Pro Ser Gly Ser  
 1 5 10 15  
 Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser  
 20 25 30  
 Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr Pro Leu  
 35 40 45  
 Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn Val Thr  
 50 55 60  
 Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala  
 65 70 75 80  
 Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly  
 85 90 95  
 Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser  
 100 105 110  
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys  
 115 120 125  
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr  
 130 135 140  
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro  
 145 150 155 160  
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr  
 165 170 175  
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly  
 180 185 190  
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly  
 195 200 205  
 Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu  
 210 215 220  
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala  
 225 230 235 240  
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn  
 245 250 255  
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe  
 260 265 270  
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro  
 275 280 285  
 Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Gln Tyr Ala Ile Val Asn  
 290 295 300  
 Asn Arg Gln Lys Asp Ala Ala Thr Ala His Thr Leu Gln Ala Pro Leu  
 305 310 315 320  
 Thr Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asn Thr Val  
 325 330 335 340  
 Phe Gln Pro Leu Ala Pro Ala Val Val Lys Leu Ser Asn Ala Leu  
 345 350 355  
 Ile Ala Thr Ile Ser Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly  
 360 365  
 Gly Ser Gly Gly Ser Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr  
 370 375 380  
 Ala Ala Ala Thr Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro  
 385 390 395 400

Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser  
 435 440 445  
 Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp  
 450 455 460  
 Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro  
 465 470 475 480  
 Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg  
 485 490 495  
 Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala  
 500 505 510  
 Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro  
 515 520 525  
 Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val  
 530 535 540  
 Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys  
 545 550 555 560  
 Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn  
 565 570 575  
 Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly  
 580 585 590  
 Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu  
 595 600 605  
 Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro  
 610 615 620  
 Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr  
 625 630 635 640  
 Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala  
 645 650

# PATENT COOPERATION TREATY

## PCT

### DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a) and Rule 39)

Applicant's or agent's file reference 9532-023-228	<b>IMPORTANT DECLARATION</b>	Date of mailing (day/month/year) <b>22 JUNE 1999</b>
International application No. PCI/US99/03265	International filing date (day/month/year) 17 FEBRUARY 1999	(Earliest) Priority Date (day/month/year) 18 FEBRUARY 1998
International Patent Classification (IPC) or both national classification and IPC Please See Continuation Sheet.		
Applicant CORIXA CORPORATION		

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below

1. ☐ The subject matter of the international application relates to:
  - a. ☐ scientific theories.
  - b. ☐ mathematical theories.
  - c. ☐ plant varieties.
  - d. ☐ animal varieties.
  - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
  - f. ☐ schemes, rules or methods of doing business.
  - g. ☐ schemes, rules or methods of performing purely mental acts.
  - h. ☐ schemes, rules or methods of playing games.
  - i. ☐ methods for treatment of the human body by surgery or therapy.
  - j. ☐ methods for treatment of the animal body by surgery or therapy.
  - k. ☐ diagnostic methods practiced on the human or animal body.
  - l. ☐ mere presentations of information.
  - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
 

☐ the description
☐ the claims
☐ the drawings
3. ☒ The failure of the nucleotide and/or amino acid sequence listing to comply with the prescribed requirements prevents a meaningful search from being carried out.
 

☒ It does not comply with the prescribed standard
 ☐ It is not in the prescribed machine-readable form
4. ☐ Other comments

**DECLARATION OF NON-ESTABLISHMENT OF  
INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US99/03265

The International Patent Classification (IPC) or National Classification and IPC are as listed below

IPC(5): A61K 38/00; C07K 1/00, C07K 16/00, C12Q 1/68; C12P 19/34; C07H 21/02, 21/04  
US Cl. 530/300, 350, 387.1; 435/6, 91.1, 91.2; 536/23.1, 24.3, 24.31, 24.32, 24.33